



Centro de Investigación en Matemáticas, A.C.

Área de Ciencias de la Computación

Nombre de la tesis:

Pruebas de hipótesis regularizadas para detección de
regiones de activación en imágenes fMRI

Que para obtener el grado de:

Maestro en Ciencias con Especialidad en Computación y
Matemáticas Industriales

P R E S E N T A:

Dora Elisa Alvarado Carrillo

Asesores:

Dr. José Luis Marroquín Zaleta
Dr. Oscar Susano Dalmau Cedeño

Guanajuato, Gto. a 29 de Julio de 2016

Integrantes del Jurado:

Presidente: Dr. Mariano José Juan Rivera Meraz (CIMAT)

Secretario: Dr. Johan Jozef Lode Van Horebeek (CIMAT)

Vocal y Co-Director de la Tesis: Dr. José Luis Marroquín Zaleta (CIMAT)

Lector Especial y Co-Director de la Tesis: Dr. Oscar Susano Dalmau Cedeño (CIMAT)

Director de la Tesis:

Dr. José Luis Marroquín Zaleta

Co-Director de la Tesis:

Dr. Oscar Susano Dalmau Cedeño

Agradecimientos

Agradezco a mis asesores, el Dr. José Luis Marroquín y el Dr. Oscar S. Dalmau, por darme la oportunidad de trabajar con ellos, por su apoyo y paciencia durante todo el proceso. También agradezco a los doctores Mariano J. J. Rivera y Johan V. Horebeek por tomarse el tiempo para leer mi trabajo y contribuir con comentarios que ayudaron a mejorar este documento.

Asimismo, agradezco al resto de los doctores con los que tomé clases, por sus enseñanzas y por siempre estar dispuestos a ayudar. Agradezco también a mis compañeros de maestría y demás amigos que hice en Guanajuato, su compañía hizo que mi estancia aquí fuera más agradable.

Agradezco a mi familia, por respaldar siempre mis decisiones y alentarme en los momentos difíciles.

Finalmente, agradezco al Consejo Nacional de Ciencia y Tecnología por el apoyo económico brindado para subsidiar mis estudios de maestría.

Índice general

Agradecimientos	3
1. Introducción	6
1.1. Pruebas de hipótesis para detección de regiones de activación en neuroimágenes	7
1.2. Estado del arte	9
1.3. Relación con el problema de segmentación	11
2. Método de pruebas de hipótesis regularizadas (RHT)	13
2.1. Formulación	13
2.2. Cálculo de la verosimilitud	14
2.2.1. Modelo de ruido blanco Gaussiano aditivo	15
2.2.2. Modelo que considera correlación espacial en los campos de ruido	16
2.3. Optimización de la función de energía $U(b)$	17
2.4. Mapeo de distribución del estadígrafo a una distribución normal . . .	18
2.5. Selección de hiperparámetros	21
2.5.1. Estimación de parámetro ν	22
2.5.2. Calibración del nivel de activación a_1	30
2.5.3. Calibración de parámetro de regularización λ	34
2.6. Recapitulación	37
3. Experimentos y resultados	39
3.1. Pruebas con la selección parámetros	39
3.1.1. Calibración de λ	39
3.2. Comparación de desempeño	42

3.3. Pruebas con datos fMRI	46
3.3.1. Datos fMRI de activación auditiva	46
3.3.2. Estimación de la distribución de ν	46
3.3.3. Variación de la constante ϵ de control de <i>FPR</i> (o mapa de p-valores generalizados)	49
3.3.4. Reducción del número de bloques	50
4. Conclusiones	53
A. Cálculo del gradiente para la funcional en la ec. (2.14)	56
A.0.5. Simplificaciones	58
B. Pseudoverosimilitud de ν	60
B.0.6. Solución	64
Bibliografía	65

Capítulo 1

Introducción

En el presente trabajo se aborda el problema de la detección de regiones de activación en imágenes de resonancia magnética funcional y se propone utilizar pruebas de hipótesis bajo un enfoque de regularización para modelar la correlación espacial que existe en las regiones de activación, considerando también en el modelo la correlación del ruido de la señal analizada.

Antes de iniciar, se presenta una breve descripción de la estructura del documento. En este capítulo se describen brevemente los conceptos básicos involucrados en el problema de la detección de activaciones en fMRI usando pruebas de hipótesis y se presenta una revisión de algunos de los métodos existentes y las estrategias que emplean para lidiar con el problema de las comparaciones múltiples.

En el capítulo 2 se habla acerca de la formulación del método propuesto (método de pruebas de hipótesis regularizadas, RHT) planteando el problema de detección de regiones de activación como un problema de segmentación. Dos modelos de ruido son considerados, el caso de ruido Gaussiano blanco aditivo y el caso que considera correlación espacial en los campos de ruido. En este mismo capítulo se propone un procedimiento de mapeo de la distribución del estadígrafo, cualquiera que sea, a una distribución normal estándar como una manera de simplificar cálculos y volver universal la tarea de selección de parámetros, la cual también se explica a detalle en el capítulo 2.

En el capítulo 3 se exponen varios experimentos que permiten comprender de mejor manera el comportamiento de los parámetros involucrados en el método RHT,

así como el desempeño del mismo al utilizar datos sintéticos y reales bajo diversas condiciones. Para el caso de los datos sintéticos, se utilizan campos generados ya sea utilizando Gibbs Sampler (más adelante se explica el procedimiento a detalle) o bien campos cuya correlación se debe a que se aplica un filtro Gaussiano. En el caso de los datos reales se ha utilizado un conjunto de volúmenes fMRI de activación auditiva, liberado por el University College de Londres.

En el capítulo 4 se discuten los resultados obtenidos, se exponen las conclusiones alcanzadas con esta investigación y el trabajo a futuro que podría realizarse en este sentido.

1.1. Pruebas de hipótesis para detección de regiones de activación en neuroimágenes

La neurociencia es un área que tradicionalmente hace uso de inferencia estadística para la interpretación de actividad cerebral. En particular, para determinar el área del cerebro responsable de una cierta tarea cognitiva, se diseña un experimento que involucre periodos de estímulo y descanso sobre uno o varios individuos y, a través de pruebas de hipótesis sobre datos tales como PET (Positron Emission Tomography) o fMRI (functional Magnetic Resonance Imaging) adquiridos, son identificadas las regiones de voxels con un grado de activación significativo. En esta última etapa se evalúan los datos obtenidos del experimento para cada voxel, de manera que se pueda encontrar evidencia estadística que apoye o no cierta hipótesis. Para ello, generalmente se asume que los datos obtenidos del experimento tienen un comportamiento lineal y se estiman los parámetros de un modelo (usualmente el modelo generalizado lineal) que se ajuste lo mejor posible a los datos. Ya que este proceso se realiza de manera individual para cada voxel, el resultado de este procedimiento es una imagen por cada parámetro estimado. Dichas imágenes son sometidas a pruebas estadísticas (usualmente paramétricas) para finalmente determinar si representan evidencia en contra de la hipótesis nula formulada, cuya definición se detalla a continuación.

Dado un conjunto \mathcal{L} de elementos (en el caso de imágenes funcionales \mathcal{L} es el conjunto de voxels), se tiene un estadígrafo $T(u)$ definido para cada elemento $u \in \mathcal{L}$ y una hipótesis nula H_0 que afirma que todos los elementos $T(u)$ son generados por

la distribución nula $P_0(T)$. Conviene plantear H_0 en función de la intersección de un conjunto de hipótesis más simples:

$$H_0 = \bigcap_{u \in \mathcal{L}} H_{0u} \quad (1.1)$$

Donde cada H_{0u} corresponde a una hipótesis nula marginal acerca de la distribución de probabilidad de las observaciones hechas en el elemento (voxel) u , es decir que $H_{0u} : T(u)$ es generada por P_0 .

La región de activación A es el conjunto de elementos u donde H_{0u} es rechazada, y se busca encontrar una estimación \hat{A} para el conjunto A . El desempeño del método elegido para obtener \hat{A} generalmente se evalúa a través de dos medidas estadísticas, la especificidad y la sensibilidad.

Sea A^c el complemento del conjunto A y $|A|$ su cardinalidad, la especificidad o tasa de falsos positivos (FPR, False Positive Rate) se define como:

$$FPR = E \left[\frac{|A^c \cap \hat{A}|}{|A^c|} \right] \quad (1.2)$$

En este sentido, conviene también definir la tasa de falsos positivos fuera de los bordes (o frontera) de la región de activación:

$$FPR_1 = E \left[\frac{|(DA)^c \cap \hat{A}|}{|(DA)^c|} \right] \quad (1.3)$$

con

$$DA = \{u : \min_{v \in A} \|u - v\| \leq 1\} \quad (1.4)$$

Por otro lado, la sensibilidad o tasa de verdaderos positivos (TPR, True Positive Rate) se define como:

$$TPR = E \left[\frac{|A \cap \hat{A}|}{|A|} \right] \quad (1.5)$$

Nótese que, debido a que en el campo de la neurociencia se considera de manera se-

parada a cada voxel y se realizan cientos de miles de comparaciones al mismo tiempo, las pruebas estadísticas agregan al problema de estimación de la región de activación A una nueva dificultad, el conocido problema de las comparaciones múltiples. Éste último se presenta cuando se realizan múltiples pruebas estadísticas independientes de manera simultánea, lo cual resulta en un incremento en el porcentaje de falsos positivos (errores tipo I) si no se realizan correcciones mayores. Es decir que de no ser ajustado el nivel de significancia (cuyo valor en muchas otras áreas suele utilizarse en 0.05) las regiones de activación detectadas contendrían un alto número de errores, el lector puede encontrar mayores detalles acerca del problema de múltiples comparaciones en neurociencia en [Berntson y Cacioppo \(2009\)](#); [Ramírez \(2011\)](#).

1.2. Estado del arte

A fin de mantener un control sobre la cantidad de falsos positivos obtenidos dado el problema de múltiples comparaciones, el procedimiento típico es declarar a un voxel como activo si el p -valor (pv) de la prueba estadística es menor a un determinado umbral, es decir que la región de rechazo se compone por:

$$\hat{A} = \{u : pv(u) \leq c\} \quad (1.6)$$

Los métodos que utilizan la técnica de umbralización son llamados métodos puntuales (PW), y se distinguen entre sí por la manera en que determinan el umbral a utilizar, entre los más conocidos se encuentran las correcciones de Bonferroni y Sidak ([Abdi, 2007](#); [Hochberg y Tamhane, 2009](#)), donde el propósito es controlar el Family-Wise Error Rate (FWER), el cual se define como sigue. Sea V el número de falsos positivos:

$$FWER = Pr(V \geq 1) = Pr(\hat{A} \neq \emptyset | A = \emptyset) \quad (1.7)$$

Por lo que al mantener $FWER \leq \alpha$, se controla la probabilidad de obtener uno o más errores de tipo I en n pruebas de hipótesis, sin importar cuantas de ellas fueron rechazadas. Las correcciones Bonferroni y Sidak, así como otras técnicas que utilizan el FWER para el análisis de neuroimágenes, son analizadas más profundamente por [Nichols y Hayasaka \(2003\)](#).

Otro enfoque es el introducido por Benjamini y Hochberg (1995), en éste se busca controlar el False Discovery Rate (FDR), es decir la proporción esperada de hipótesis nulas que fueron rechazadas de manera errónea. Sea V el número de falsos positivos y R el número de hipótesis nulas rechazadas:

$$FDR = E \left[\frac{V}{R} \right] = E \left[\frac{|A^c \cap \hat{A}|}{|\hat{A}|} \right] \quad (1.8)$$

El procedimiento considera todos los p-valores antes de decidir si rechazar o aceptar las hipótesis, para ello ordena de forma creciente los p-valores resultantes de las n pruebas de hipótesis, es decir $pv_1 \leq pv_2 \leq \dots \leq pv_n$, y encuentra el índice k tal que k sea el mayor i para el cual $pv_i \leq \frac{i}{n}\alpha$, luego las hipótesis asociadas a los índices $1, 2, \dots, k$ son rechazadas, y el FDR es controlado a un nivel α . Una de las razones por la que estas estrategias son poco recomendadas en estudios de PET y fMRI es que los umbrales obtenidos son tan altos que incrementan los falsos negativos (errores tipo II), es decir disminuyen la potencia estadística. Debido a esto, más recientemente se han presentado propuestas cuyo interés es, además de mantener bajo control el FPR, maximizar la sensibilidad (TPR) aprovechando el hecho de que las regiones de activación son por lo general grupos de voxeles (clusters) conectados, y no elementos aislados como se considera en los métodos discutidos anteriormente.

Propuestas como la de Worsley et al. (1992), Kiebel et al. (1999), Hayasaka y Nichols (2003) abordan el problema mediante la teoría de Campos Aleatorios Gaussianos, bajo la suposición de que la correlación espacial de los datos es conocida o puede ser estimada, o bien que los datos siguen una cierta distribución, lo cual en la práctica no se cumple (Purdon y Weisskoff, 1998; Zarahn et al., 1997), por lo que generalmente se realiza un procesamiento previo de suavizado de los datos para asegurar que dichas suposiciones se cumplan, con lo que se pierde resolución espacial.

Un enfoque distinto se aborda en Friston et al. (1994), donde se propone que la detección de activaciones no dependa únicamente de la intensidad del estadígrafo T , sino que esta decisión involucre la extensión espacial de los grupos de voxeles conectados resultantes al umbralizar T , de modo que se tome en cuenta el grado de conectividad, procedimientos similares son presentados por Bullmore et al. (1999) y Zhang et al. (2009), en cuyo caso se plantea la idea de calcular la distribución

empírica de la masa del grupo de voxels conectados, que es la suma de diferencias de las intensidades que pertenecen al mismo cluster menos la intensidad del umbral. Este tipo de métodos han probado mejorar la sensibilidad con respecto a aquellos que no consideran información de conectividad, pues detectan regiones con amplia extensión espacial incluso con activaciones moderadas. No obstante, los métodos fallan para detección de regiones de tamaños pequeños, y dado que parten de la umbralización del estadígrafo T para calcular un nuevo estadígrafo que involucre el grado de conexión de las regiones de activación, una debilidad importante es la fuerte dependencia del resultado final a la selección del umbral, cuyo valor es elegido arbitrariamente.

Smith y Nichols (2009) abordan el problema de selección de umbral mediante una metodología que pretende mejorar las regiones que presentan contigüidad espacial, como lo hacen los métodos basados en la detección de clusters a través de umbrales, pero sin depender de un solo umbral, en cambio calculan un estadígrafo a través de la suma ponderada de varios tamaños de cluster obtenidos con umbrales diferentes, lo cual produce resultados en los que la diferencia entre ruido y región de activación es más evidente que en otros métodos basados en clusters. Otro método libre de umbrales es el presentado por Marroquin et al. (2012), en el cual se utilizan operadores morfológicos de erosión y dilatación con un conjunto de elementos estructurantes de diversos tamaños, basados en las ideas de Haralick et al. (1987) y Serra (1984), lo que permite detectar regiones con niveles de activación moderados y extensión espacial amplia, es decir que los resultados son comparables a los obtenidos con métodos basados en la detección de clusters a través de umbrales.

1.3. Relación con el problema de segmentación

Una manera de interpretar el problema de encontrar regiones de activación sobre neuroimágenes es abordarlo desde el punto de vista de la segmentación de imágenes. Considérese que se tiene el conjunto \mathcal{L} de elementos, y como ya se ha mencionado en la sección 1.1, se desea encontrar una estimación de la región A , que corresponde al conjunto de elementos donde H_{0u} es rechazada. Se puede también ver a \mathcal{L} como un conjunto de regiones A_0, A_1, \dots, A_k que no se superponen, cuyos niveles de activación

a_0, a_1, \dots, a_k son distintos entre sí y $a_0 = 0$, es decir que la región de activación se define como $A = \bigcup_{k>0} A_k$ y A_0 corresponde a la región de no activación A^c . Uno puede definir una función indicadora b_k de la región A_k , el problema entonces consiste en encontrar el campo $b(x) = (b_1(x), b_2(x), \dots, b_k(x))$, el cual puede ser modelado usando teoría Campos Aleatorios Markovianos (MRF, por sus siglas en inglés). Utilizar este enfoque tiene la particularidad de que se incluyen restricciones de correlación espacial sobre la solución, por lo que conviene en el caso de la detección de regiones de activación sobre fMRI. En este sentido, una propuesta eficiente es presentada por Rivera et al. (2007), en donde se plantea relajar el problema y en vez de encontrar el campo $b(x)$ (es decir, obtener un mapa de etiquetas) proponen calcular la probabilidad de que el dato observado en un pixel corresponda a una determinada región, esto a través de MRFs con potenciales cuadráticos de entropía controlada, obtenidos utilizando información a priori acerca de la segmentación.

En el capítulo siguiente se retoma la relación descrita entre el problema de segmentación y la detección de regiones de activación en fMRI, así como la estrategia de utilizar campos aleatorios markovianos para dar solución al problema, pues son piezas importantes para describir el enfoque de regularización propuesto en este documento.

Capítulo 2

Método de pruebas de hipótesis regularizadas (RHT)

En este capítulo se describe la formulación del método de pruebas de hipótesis regularizadas para la detección de regiones de activación en imágenes fMRI. Asimismo, se describe el procedimiento seguido para la estimación de los hiperparámetros involucrados en el método propuesto.

2.1. Formulación

El método propuesto aborda la detección de regiones de activación como un problema de segmentación. Como se explica al final del capítulo 1, uno puede hacer la formulación como sigue: Dado un conjunto de elementos \mathcal{L} , el problema consiste en particionar \mathcal{L} en regiones A_0, \dots, A_K que no se superpongan, de modo que $\mathcal{L} = \bigcup_k A_k$, $A_i \cap A_j = \emptyset$ para $i \neq j$, y donde $A = \bigcup_{k>0} A_k$ y A_0 representa la región de no activación A^c . Se busca un campo $b(u)$ compuesto de funciones indicadoras de cada región $b_k(u)$, con $u \in \mathcal{L}$, es decir que $b_k(u) = 1$ ssi $u \in A_k$ y $b_k(u) = 0$ en otro caso, por lo cual se tienen las restricciones siguientes:

$$b_k(u) \in \{0, 1\}, \forall k, u \quad (2.1)$$

$$\sum_{k=0}^K b_k(u) = 1, \forall u \quad (2.2)$$

Se puede utilizar un enfoque Bayesiano y modelar al campo $b(u)$ como un campo aleatorio Markoviano, de modo que la distribución a posteriori se puede escribir como:

$$P(b|\hat{T}) = \frac{1}{Z} \exp(-U(b)) \quad (2.3)$$

con:

$$U(b) = \sum_u \sum_k (-\log l_k(u)) b_k^2(u) + \lambda \sum_{\langle u,v \rangle} V(b_k(u), b_k(v)) \quad (2.4)$$

Donde $l_k(u)$ representa la verosimilitud del elemento u de pertenecer a la región A_k , $V(b(u), b(v))$ corresponde al potencial de Ising:

$$V(b(u), b(v)) = \sum_k (b_k(u) - b_k(v))^2 \quad (2.5)$$

y la notación $\langle u, v \rangle$ indica que esto se calcula para todos los vecinos de u tales que $\{v \in \mathcal{L} : |u - v| = 1\}$. Aunque se puede encontrar la estimación máxima a posteriori minimizando la función (2.4) sujeto a las restricciones (2.1) y (2.2), se trata de un problema de optimización combinatoria y encontrar una solución es en general difícil, una estrategia más eficiente es relajar la restricción (2.1) como sigue:

$$b_k(u) \geq 0, \forall k, u \quad (2.6)$$

Al aplicar esta relajación a la restricción (2.1), $b_k(u)$ puede ahora interpretarse como una aproximación de la distribución marginal a posteriori de la función indicadora de cada región A_k .

2.2. Cálculo de la verosimilitud

Asumiendo que se tiene un modelo para generar campos del estadígrafo T bajo la hipótesis nula, uno puede estimar la distribución nula y mapear los campos T (ver sección 2.4) a campos con una distribución normal estándar \hat{T} , a fin de facilitar el cálculo de los campos de verosimilitud. Sea $P_0(\hat{T}(u)) \stackrel{def}{=} P(\hat{T}(u)|H_0)$, la verosimilitud

l_k es:

$$l_k(u) = P(\hat{T}(u)|u \in A_k) = P_0(\hat{T}(u) - a_k) \quad (2.7)$$

Con $k \in \{0, 1\}$, de modo que $A = A_1$ representa la región de activación. En las siguientes subsecciones se presenta el desarrollo de la expresión (2.7) para dos casos: uno usando un modelo de ruido blanco Gaussiano aditivo, y el otro usando un modelo que considera correlación espacial en los campos de ruido.

2.2.1. Modelo de ruido blanco Gaussiano aditivo

Bajo un modelo de ruido Gaussiano aditivo, se tiene que la verosimilitud expresada en la ecuación (2.7) equivale a:

$$\begin{aligned} l_k(u) &= P_0(\hat{T}(u) - a_k) \\ &= \frac{1}{\sqrt{2\pi}} \exp[-(\hat{T}(u) - a_k)^2/2]. \end{aligned} \quad (2.8)$$

Para este modelo de ruido, se tiene la siguiente función de energía:

$$U(b) = \frac{1}{2} \sum_u \sum_k \left(\hat{T}(u) - a_k \right)^2 b_k^2(u) + \lambda \sum_{\langle u,v \rangle} V(b(u), b(v)). \quad (2.9)$$

donde $V(x, y) = (x - y)^2$ y (2.9) sujeta a las restricciones (2.2) y (2.6). Este enfoque es muy similar al considerado en Rivera et al. (2007), texto al que el lector se puede dirigir para mayores detalles acerca de la derivación de la fórmula anterior. Nótese que el primer término en la función de energía se origina de la verosimilitud en (2.8), mientras que el segundo corresponde al término de regularización de la granularidad de las regiones detectadas, que se controla mediante el hiperparámetro λ .

2.2.2. Modelo que considera correlación espacial en los campos de ruido

Hasta este punto se ha trabajado bajo la suposición de un modelo de ruido blanco Gaussiano aditivo, sin embargo en la práctica esto no es completamente cierto, pues existe un grado de correlación en los campos de ruido, lo que afecta el desempeño de todos los métodos. Un planteamiento distinto al utilizado hasta ahora consiste en asumir que sí existe correlación en los campos de ruido n , en este caso la distribución de $T(u)$ dado que pertenece a la clase $C(u)$:

$$P(T(u)|C(u)) = P(n(u) = T(u) - a_{C(u)}) = \frac{1}{Z} \exp[-U(n)], \quad (2.10)$$

con $n = T(u) - a_{C(u)}$ y donde:

$$U(n) = \frac{1}{2} \sum_u n(u)^2 + \nu \sum_{\langle u,v \rangle} (n(u) - n(v))^2 \quad (2.11)$$

$$= \frac{1}{2} \sum_u (T(u) - a_{C(u)})^2 + \nu \sum_{\langle u,v \rangle} V(T(u) - a_{C(u)}, T(v) - a_{C(v)}) \quad (2.12)$$

Así que para este modelo de ruido la verosimilitud se escribe como sigue:

$$\begin{aligned} l_C(u) &= P_0(\hat{T}(u) - a_{C(u)}) \\ &= \frac{1}{Z} \exp \left[- \left[\frac{1}{2} (\hat{T}(u) - a_{C(u)})^2 + \nu \sum_{\langle u,v \rangle} V(\hat{T}(u) - a_{C(u)}, \hat{T}(v) - a_{C(v)}) \right] \right] \end{aligned} \quad (2.13)$$

donde Z es una constante que depende de ν y a_k , $V(x, y) = (x - y)^2$, nótese también que cuando $\nu = 0$ uno obtiene la expresión 2.7. En otro caso, para $\nu \neq 0$, se obtiene la siguiente expresión como función de energía:

$$\begin{aligned} U(b) &= \frac{1}{2} \sum_u (\hat{T}(u) - a_{C(u)})^2 b_C^2(u) \\ &\quad + \nu \sum_{\langle u,v \rangle} V(\hat{T}(u) - a_{C(u)}, \hat{T}(v) - a_{C(v)}) b_C(u) b_C(v), \end{aligned}$$

suponiendo que se cumplen las restricciones (2.1) y (2.2) sobre b , ahora relajando la

restricción (2.1), usando en cambio (2.6) e incluyendo la condición de suavidad sobre b , se obtiene:

$$\begin{aligned}
&= \frac{1}{2} \sum_u \sum_k (\hat{T}(u) - a_k)^2 b_k^2(u) + \lambda \sum_{\langle u,v \rangle} V(b_k(u), b_k(v)) \\
&\quad + \nu \sum_{\langle u,v \rangle} \sum_{k,j} V(\hat{T}(u) - a_k, \hat{T}(v) - a_j) b_k(u) b_j(v), \tag{2.14}
\end{aligned}$$

donde (2.14) está sujeta a las restricciones (2.2) y (2.6), las cuales hacen que se obtenga una aproximación a la distribución a posteriori de la función indicadora de cada región A_k . Nótese que el primer y tercer término en la ecuación (2.14) provienen de la verosimilitud en (2.13), mientras que el segundo corresponde al término de regularización de la granularidad de las regiones, al igual que en el caso anterior donde se supone un modelo de ruido blanco Gaussiano aditivo.

2.3. Optimización de la función de energía $U(b)$

Debido a que en la práctica interesa únicamente determinar si un elemento se encuentra activo o no, sin hacer distinciones de acuerdo a su nivel de activación, en este trabajo se estudia el caso cuando $K = 1$. Nótese que uno puede simplificar $U(b)$ tomando $K = 1$ (es decir, considerar dos clases, activación y no activación) y $a_0 = 0$, sin que eso afecte el desempeño del método, pues experimentalmente se ha probado usando $K > 1$ y los resultados obtenidos son muy similares con la desventaja de que la complejidad es mayor.

Si se usara un modelo de ruido blanco Gaussiano aditivo, la optimización de la función (2.9) podría realizarse mediante un procedimiento como Gauss-Seidel, dado que la matriz Hessiana es definida positiva. Sin embargo, una manera sencilla de optimizar es utilizar descenso con gradiente simple. El procedimiento del cálculo del gradiente se puede revisar en el apéndice A.

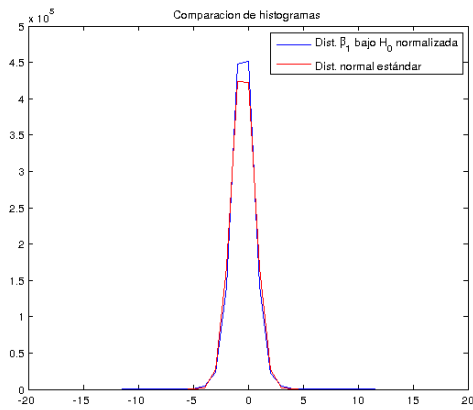
2.4. Mapeo de distribución del estadígrafo a una distribución normal

Nótese que la formulación propuesta asume un modelo de ruido aditivo Gaussiano, suposición que no siempre se cumple, por ejemplo en la Figura 2.1 se ha realizado una comparación de la distribución bajo H_0 de dos estadígrafos: el coeficiente β_1 del modelo lineal generalizado (considerando que se inicia en β_0) y el estadígrafo F. Ambos estadígrafos fueron calculados a partir de datos reales de fMRI, los datos han sido normalizados en ambos casos (su media es cero y su varianza uno), sin embargo al observar los histogramas y los QQPlot resultantes comparados contra la distribución normal estándar las diferencias son notorias, principalmente al utilizar el estadígrafo F.

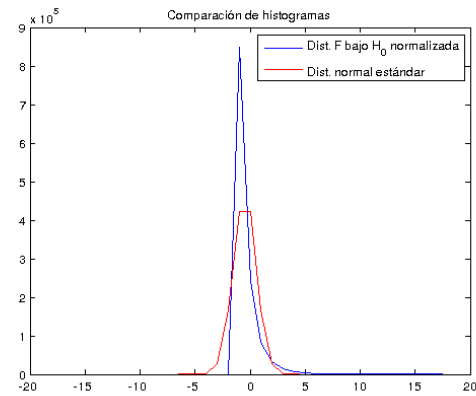
Por esta razón, se propone realizar una transformación de la distribución del estadígrafo T bajo H_0 a una distribución normal, además de que esto aporta a la generalización del procedimiento de selección adecuada de parámetros. Para realizar esta transformación, se pueden utilizar las siguientes dos propiedades para :

1. Si x es una variable aleatoria y F es su función de densidad acumulada (cdf), entonces la distribución de la variable $F(x)$ es uniforme.
2. Si u es una variable aleatoria uniforme y G la cdf de la distribución Gaussiana, entonces la función de distribución para la variable aleatoria $G^{-1}(u)$ será Gaussiana.

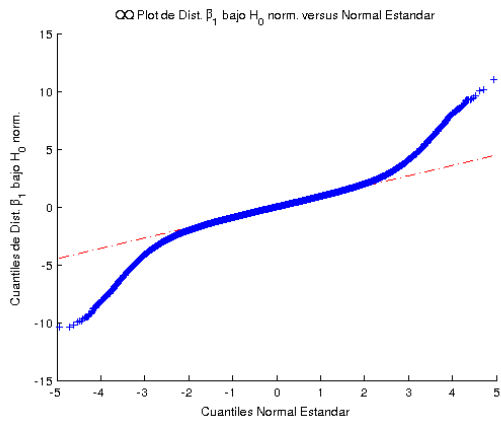
Es decir que para cualquier variable aleatoria x , la variable $G^{-1}(F(x))$ sigue una distribución Gaussiana. De acuerdo con lo anterior, si uno conoce la función de densidad acumulada de x o tiene manera de estimarla obteniendo la función de densidad acumulada empírica (ecdf) $\hat{F}(x)$, se puede aplicar una transformación a la distribución de la variable x para mapearla a una distribución Gaussiana. De la misma manera, la distribución del estadígrafo T bajo H_0 puede ser transformada a una distribución Gaussiana de media $\mu = 0$ y desviación estándar $\sigma = 1$, una manera de realizar esta transformación se describe en el algoritmo 1, donde se utiliza la función error inversa (erf^{-1}) aprovechando el hecho de que la función error (erf) es casi idéntica a la función de distribución Gaussiana, con una única diferencia de



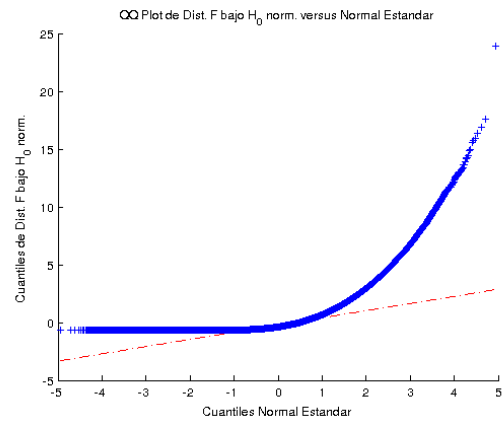
(a)



(b)



(c)



(d)

Figura 2.1: Arriba: Comparación de histograma del campo T con el histograma de un conjunto de valores aleatorios con una distribución normal estándar, usando (a) β_1 del modelo lineal generalizado y (b) usando el estadígrafo F . Abajo: QQPlot de la Distribución normalizada del estadígrafo bajo H_0 , usando (c) el β_1 del modelo lineal generalizado y usando (d) el estadígrafo F .

escala y traslación. La función error (o error de gauss) es una función especial que se define como sigue :

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{2.15}$$

El nombre *función error* proviene de la relación con la función de distribución acu-

umulativa de una variable normal estándar (Φ), pues se tiene que:

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \quad (2.16)$$

Si uno tiene un número complejo z , no existe un único número complejo w que satisfaga $\operatorname{erf}(w) = z$, por lo que la función inversa podría tener múltiples valores de salida para un elemento de entrada. No obstante, para $-1 < x < 1$ si existe un único número real, definido $\operatorname{erf}^{-1}(x)$ con erf^{-1} como la función error inversa, que satisface $\operatorname{erf}(\operatorname{erf}^{-1}(x)) = x$. Por esta razón, es común que en muchas herramientas se defina a la función erf^{-1} solamente en el dominio $(-1, 1)$. Para más detalles acerca de la función error y la función error inversa, consultar [Abramowitz y Stegun \(1972\)](#).

Por otro lado, retomando la ecuación (2.16):

$$\Phi(x) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \quad (2.17)$$

$$\Phi(x) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (2.18)$$

$$2\Phi(x) = 1 + \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \quad (2.19)$$

$$2\Phi(x) - 1 = \operatorname{erf} \left(\frac{x}{\sqrt{2}} \right) \quad (2.20)$$

$$\operatorname{erf}^{-1}(2\Phi(x) - 1) = \frac{x}{\sqrt{2}} \quad (2.21)$$

$$x = \sqrt{2} \operatorname{erf}^{-1}(2\Phi(x) - 1) \quad (2.22)$$

o bien,

$$x = \sqrt{2} \operatorname{erf}^{-1}(\theta), \quad (2.23)$$

con $\theta = 2\Phi(x) - 1$ y donde la variable x sigue una distribución normal estándar, la única restricción en el desarrollo anterior está sobre $\Phi(x)$, que sólo toma valores entre cero y uno, pues se trata de una función de distribución acumulada. De acuerdo con las propiedades explicadas al inicio de este apartado, uno podría sustituir $\Phi(x)$ por una variable aleatoria uniformemente distribuida en el intervalo $[0,1]$ y se seguiría

cumpliendo que $x \sim \mathcal{N}(0, 1)$. Por ejemplo, si se quisiera generar números aleatorios con una distribución normal estándar a partir de números aleatorios con distribución uniforme, se usaría $x = \sqrt{2}erf^{-1}(u)$, con $u \in [-1, 1]$, que es en realidad el rango en el que se encuentra la expresión (2.23).

En el algoritmo 1, primero se genera un número grande n de muestras del estadígrafo T bajo H_0 (T_0) para obtener su distribución, los valores de todas las muestras se almacenan en un vector w y son ordenados en orden ascendente, llámese N al tamaño de w . La idea es la siguiente, si uno tiene una muestra del estadígrafo T y quiere mapear todos sus valores a una distribución normal estándar, toma cada valor $T(u)$, $u \in \mathcal{L}$ y recorre los valores en el vector w hasta ubicar la mínima posición j del vector en la que se cumple que $w(j) > T(u)$, y se llama a esta posición k (ver ecuación (2.24)). Ahora bien, k/N representa la probabilidad acumulada del valor $T(u)$, al ser una probabilidad su valor se encuentra en el intervalo $[0, 1]$, es decir se usa en la ecuación (2.23) a $\theta = 2(k/N) - 1$, sin embargo hay que considerar que si $k = 0$ o $k = N$ el valor θ obtenido queda fuera del intervalo $(-1, 1)$ y debe proyectarse a un valor dentro del intervalo para poder utilizar la función $erf^{-1}(\theta)$ y transformar el valor $T(u)$ al equivalente bajo la distribución normal estándar $T(\hat{u})$, usando la expresión (2.23).

La Figura 2.2 muestra las gráficas QQPlot de la distribución del estadígrafo bajo H_0 una vez que ha sido mapeada, utilizando nuevamente como estadígrafos el coeficiente β_1 del modelo generalizado lineal y el estadígrafo F , en ambos casos la distribución del estadígrafo muestra gran similitud con la de los datos con distribución normal estándar.

2.5. Selección de hiperparámetros

La expresión (2.14) está compuesta de hiperparámetros cuyo valor impacta sobre el desempeño del método principalmente en los siguientes aspectos: el valor para la región de activación a_1 , que permite controlar el FPR; el término de regularización λ , influye en la granularidad de las regiones detectadas como activas, mayores valores sobre este parámetro provocan resultados más suaves; el término ν , que está relacionado con el grado de correlación en los campos de ruido. A continuación se describen

Algoritmo 1 Mapeo de la distribución T bajo H_0 a una distribución normal estándar

Input: M_0 , el modelo que permita generar muestras del estadígrafo T bajo H_0 . T , el estadígrafo cuya distribución se desea mapear a una normal estándar.

Output: \hat{T} , el estadígrafo T mapeado a una distribución normal estándar

- 1: Generar n grande de muestras del estadígrafo T bajo H_0 (T_0), almacenar los valores de las muestras en el vector w .
- 2: Ordenar los valores en w de manera ascendente.
- 3: **for each** $T(u), u \in \mathcal{L}$, **do**
- 4: Recorrer los elementos del vector w hasta encontrar el primer elemento mayor a $T(u)$, guardar la posición de este elemento en la variable k :

$$k = \underset{j}{\text{mín}} v(j) > T(u), j \in [1, N]$$

- 5: Cambiar k a un valor en el intervalo $(-1, 1)$, por ejemplo usando:

$$\theta = 2k/N - 1$$

(Nota: si $k = 0$ o $k = N$ el valor obtenido queda fuera del intervalo $(-1, 1)$ y debe proyectarse a un valor dentro del intervalo)

- 6: Utilizar la función de error inversa para mapear al valor correspondiente de una distribución Gaussiana de media $\mu = 0$ y $\sigma = 1$:

$$\hat{T}_0(u) = \sqrt{2} \text{erf}^{-1}(\theta)$$

- 7: **end for**
-

los procedimientos propuestos para la selección de los hiperparámetros mencionados.

2.5.1. Estimación de parámetro ν

El parámetro ν controla la correlación espacial del ruido en el modelo de la ecuación (2.14). Cuando se trabaja con datos reales se sabe que existe un grado de correlación del ruido que afecta el desempeño de todos los métodos existentes hasta la fecha, un valor adecuado para ν puede mejorar notablemente el desempeño del método. El valor de este parámetro puede obtenerse mediante el cálculo de la máxima log-pseudo-verosimilitud (Besag, 1974, 1975), que se calcula a partir de la

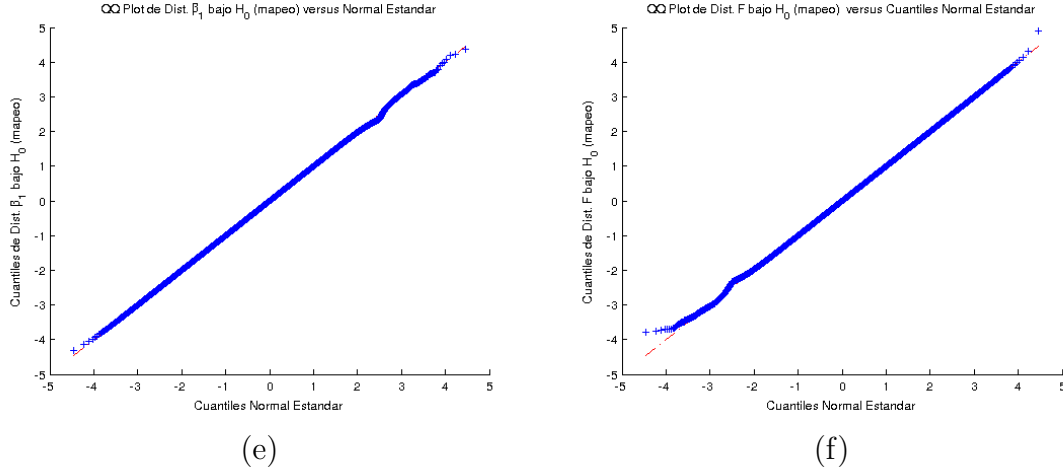


Figura 2.2: QQPlot de la Distribución mapeada del estadígrafo bajo H_0 , usando el β_1 del modelo lineal generalizado (Figura (e)) y usando el estadígrafo F (Figura (f)).

siguiente fórmula.

$$l(\nu) = \log \left(\prod_u P(n(u)|n(v), v \in N_u) \right) \quad (2.24)$$

$$= \sum_u \left(-U_u(n(u)) + B(u) - \log(\sigma(u)) - \log(\sqrt{2\pi}) \right) \quad (2.25)$$

donde:

$$P(n(u)|n(v), v \in N_u) = \frac{\exp(-U_u(n(u)))}{\int_{-\infty}^{\infty} \exp(-U_u(x)) dx} \quad (2.26)$$

y

$$U_u(x) = \frac{1}{2}x^2 + \nu \sum_{u \in N_u} (x^2 - 2xn(v) + n^2(v)) \quad (2.27)$$

Se puede mostrar (ver apéndice B) que la expresión (2.26) equivale a:

$$P(n(u)|n(v), v \in N_u) = \frac{\exp(-U_u(n(u)))}{\exp(-B(u)) \sqrt{2\pi}\sigma(u)} \quad (2.28)$$

con:

$$\mu(u) = \frac{\nu \sum_{v \in N_u} n(v)}{\frac{1}{2} + \nu |N_u|} \quad (2.29)$$

$$\sigma(u) = \frac{1}{\sqrt{2(\frac{1}{2} + \nu |N_u|)}} \quad (2.30)$$

$$B(u) = \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2 \left(\sum_{v \in N_u} n(v) \right)^2}{\frac{1}{2} + \nu |N_u|} \quad (2.31)$$

Lo que interesa maximizar es entonces:

$$l(\nu) = \log \left(\prod_u P(n(u) | n(v), v \in N_u) \right) \quad (2.32)$$

$$= \sum_u \left(-U_u(n(u)) + B(u) - \log(\sigma(u)) - \log(\sqrt{2\pi}) \right) \quad (2.33)$$

$$= \mathbf{E} \left[-\frac{1}{2} n^2(u) - 4\nu n^2(u) + 2\nu n(u) \sum_{v \in N_u} n(v) - \frac{\nu^2}{\frac{1}{2} + \nu |N_u|} \left(\sum_{v \in N_u} n(v) \right)^2 + \frac{1}{2} \log \left(\frac{1}{2} + 4\nu \right) \right] \quad (2.34)$$

Descartando los términos que no dependen de ν , la expresión a maximizar es:

$$2\nu \bar{S}_2 - 4\nu \bar{S}_1 - \frac{\nu^2}{\frac{1}{2} + 4\nu} \bar{S}_3 + \frac{1}{2} \log \left(\frac{1}{2} + 4\nu \right) \quad (2.35)$$

con:

$$S_1 = n^2(u) \quad (2.36)$$

$$S_2 = n(u) \sum_{v \in N_u} n(v) \quad (2.37)$$

$$S_3 = \left(\sum_{v \in N_u} n(v) \right)^2 \quad (2.38)$$

Para entender mejor el desarrollo anterior, considérese el experimento siguiente.

Sea la región de activación A , definida en un espacio (o rejilla) regular \mathcal{L} de 50×50 :

$$A = u \in \mathcal{L} : (u_1 - 10)^2 + (u_2 - 30)^2 \leq 16 \quad (2.39)$$

y un modelo de observación del campo T :

$$T(u) = G_\gamma * (n(u) + aI_A(u)) \quad (2.40)$$

Donde G es un kernel Gaussiano bi-dimensional con parámetro γ , $*$ denota convolución espacial y $n(u)$ con $u \in \mathcal{L}$, son variables aleatorias independientes con distribución Gaussiana estándar, I_A es la función indicadora de la región A y a el nivel de activación, en este experimento $a = 2$ y bajo la hipótesis nula $a = 0$. La correlación espacial ρ de los campos bajo H_0 fue estimada como la esperanza del producto de los pares de vecinos en el campo T , de modo que según el valor de σ utilizado se generan campos con distinta correlación espacial, experimentalmente se obtuvieron las correspondencias mostradas en la Tabla 2.1.

Se generaron 1000 campos T para $\rho = \{0.00, 0.10, 0.25, 0.50\}$, en cada caso se calculó la media de S_1, S_2, S_3 tanto para los campos mapeados a una distribución normal estándar como para los campos sin mapear, la Figura 2.3 muestra la evaluación de la log-verosimilitud de ν ($l(\nu)$) para ambos casos, es evidente que existe un cambio en el comportamiento de la función luego de que los datos son mapeados, este cambio en $l(\nu)$ puede explicarse de la siguiente manera, en la ecuación (2.38) se está considerando que el ruido tiene una cierta distribución de probabilidad, sin embargo al realizar el mapeo de los campos, éstos son transformados y la función de

Cuadro 2.1: Correlación espacial generada usando el modelo (2.40) variando γ

γ	Correlación espacial ρ
0.3682	0.0528
0.4091	0.1012
0.4409	0.1495
0.4690	0.2018
0.4957	0.2528
0.5220	0.3049
0.5490	0.3518
0.5775	0.4043
0.6085	0.4524
0.6433	0.5029

probabilidad cambia, con lo que el argumento que maximizaría $l(\nu)$ originalmente ya no corresponde con el obtenido usando los campos sin mapear.

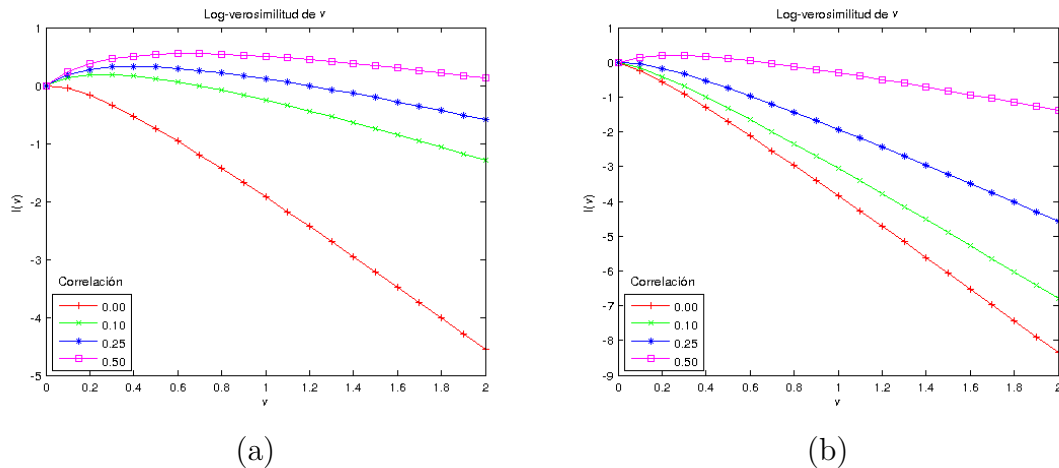


Figura 2.3: ν vs. $l(\nu)$ para distintos grados de correlación espacial $\rho = \{0.00, 0.10, 0.25, 0.50\}$ en los campos T , a la izquierda resultados sin mapear los campos, a la derecha resultados usando campos mapeados a una dist. normal estándar

El experimento se repitió utilizando el método de Gibbs Sampler (GS) para generar el campo T , usando $P(n(u)|n(v), v \in N_u)$, explicado en el algoritmo 2. La Figura 2.4 muestra que también se presenta un cambio al mapear los campos generados usando GS, nuevamente se observa que el argumento que maximiza $l(\nu)$ es alterado al realizar el mapeo.

Algoritmo 2 Gibbs Sampler

Input: Un campo inicial $n^{(0)}$ **Output:** Un campo T cuya distribución sigue el modelo de ruido (2.12)

- 1: Inicializar $[n(u)]^{(0)} \sim N(0, 1)$ para toda $\nu \in \mathcal{L}$
- 2: **for** $t = 0, 1, \dots, \text{max_}t$ **do**
- 3: **for each** $\nu \in \mathcal{L}$ **do**
- 4: Muestrear un valor para $[n(u)]^{(t+1)}$ de:

$$P(n(u)|n(v), v \in N_u) = \frac{\exp(-U_u(n(u)))}{\int_{-\infty}^{\infty} \exp(-U_u(x)) dx}$$

Note que en la expresión anterior, el denominador es en realidad una constante (ver ecuación 2.38 y apéndice B), por lo que la distribución de $P(n(u)|n(v), v \in N_u)$ en realidad es equivalente a $\exp(-U_u(n(u)))$, es decir que el nuevo valor $[n(u)]^{(t+1)} \sim N(\mu_u, \sigma_u)$, con μ_u, σ_u definidos en (2.29), (2.30).

- 5: **end for**
 - 6: $t = t + 1$
 - 7: **end for**
 - 8: Terminar con $T = n^{(\text{max_}t)}$.
-

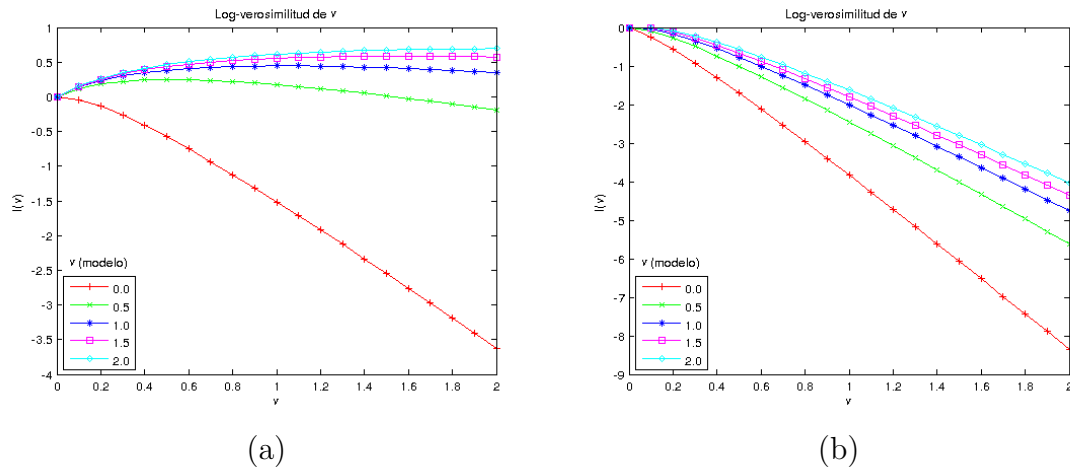


Figura 2.4: ν vs. $l(\nu)$ usando campos T generados con GS, a la izquierda resultados sin normalizar los campos, a la derecha resultados usando campos normalizados

Los resultados anteriores son indicador de que existe de forma implícita una varianza σ^2 asociada a cada valor de ν , es decir $\sigma(\nu)$, y por lo tanto al modificar dicha varianza durante el proceso de mapeo a una distribución normal estándar, la ν que se estima ya no corresponde a aquella que tenían los datos originalmente, y por lo tanto no es en realidad óptima. Si uno genera datos mediante Gibbs Sampler con el modelo (2.12), puede comprobar que la varianza de las variables aleatorias ν depende de ν :

Si se tiene un campo T generado bajo H_0 con distribuciones marginales $P_0(T(u)) \sim \mathcal{N}(0, 1)$ y se quiere estimar ν para dicho campo, se requiere encontrar un factor de escala apropiado β , de modo que para la ν estimada $\hat{\nu}$ se cumpla que la varianza (var) de $\beta T(u)$ sea igual a $\sigma_n^2(\hat{\nu})$, es decir que exista consistencia entre $\hat{\nu}$ y la varianza de $\beta T(u)$. Una manera de estimar simultáneamente β y $\hat{\nu}$ es utilizar el algoritmo EM (Dempster et al., 1977), descrito en el algoritmo 3 para el caso específico de la estimación de ν . Partiendo del hecho de que los datos se encuentran mapeados a una distribución normal estándar, el procedimiento busca la consistencia de la ν con la σ medida en los datos, mediante un procedimiento iterativo que consiste en estimar $\hat{\nu}$ para el conjunto de campos escalados $n = \beta \hat{T}$ mediante la ecuación obtenida en (2.35) (paso M), para luego realizar el cálculo de la varianza $\sigma_n^2(\hat{\nu}) = var(\hat{\nu})$ (paso E) sobre campos generados usando (2.12), por ejemplo utilizando Gibbs Sampler,

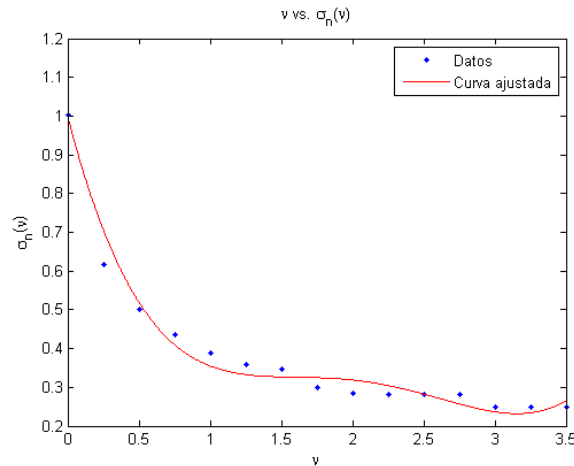


Figura 2.5: ν vs. $\sigma(\nu)$, para campos T generados mediante Gibbs Sampler usando ν modelo= $\{0.00, 0.25, \dots, 3.5\}$ (azul), y una curva ajustada ($0.04\nu^4 - 0.38\nu^3 + 1.12\nu^2 - 1.43\nu + 1$) a los datos obtenidos (rojo)

$\sigma_n^2(\hat{\nu})$ se utiliza para actualizar el factor de escala para los campos \hat{T} .

Algoritmo 3 Estimación de ν óptima (ν^*) para campos mapeados a una distribución normal estándar

Input: Un conjunto de campos T generados bajo H_0 que han sido mapeados para tener una distribución normal estándar (\hat{T})

Output: β , el factor apropiado que cumpla $var(\beta\hat{T}(u)) = \sigma_n^2(\hat{\nu})$ para el conjunto de campos \hat{T} . ν^* , el valor óptimo del parámetro ν para el conjunto de campos \hat{T}

- 1: Inicializar $\beta = 1$
 - 2: **repeat**
 - 3: Paso M: encontrar $\hat{\nu}(\beta) = \arg \max_{\nu} l(\nu)$ con $n = \beta\hat{T}$.
 - 4: Paso E: encontrar $\sigma_n^2(\hat{\nu}) = var(\hat{n}(\hat{\nu}))$, donde $n(\hat{\nu})$ es un campo generado con con (2.12) con $\nu = \hat{\nu}$, por ejemplo usando Gibbs Sampler.
 - 5: Poner $\beta = \sigma_n(\hat{\nu})$
 - 6: **until** convergencia
 - 7: Terminar con $\nu^* = \hat{\nu}$
-

Nótese que el Paso E del algoritmo puede eficientarse en la implementación, ya que no es necesario correr el Gibbs Sampler en cada caso, pues la función $\sigma_n(\hat{\nu})$ puede estimarse fuera de línea para todas las ν de interés (utilizando Gibbs Sampler). La Figura 2.5 muestra la gráfica ν vs. $\sigma_n(\nu)$ obtenida usando campos T con $\nu =$

$\{0.00, 0.25, \dots, 2\}$, como es de esperarse para campos con un parámetro ν bajo (poca correlación) la $\sigma(\nu)$ estimada es más cercana a 1.0, pero su valor va decreciendo conforme aumenta el valor de ν .

Como experimento para verificar el funcionamiento del algoritmo EM propuesto, se tomaron 1000 campos T generados con el método Gibbs Sampler usando como ν modelo $\nu \in C$, con $C = \{0.0, 0.25, \dots, 3.5\}$, los campos fueron mapeados a una distribución normal estándar. Se realizó la estimación de la ν utilizando el algoritmo 3 y se obtuvieron los valores mostrados en la tabla las estimaciones obtenidas se aproximan al valor ν modelo usado para generar los campos.

Cuadro 2.2: Estimación de parámetro ν , usando el algoritmo 3

ν modelo GS	ν estimada
0.00	0.00523
0.25	0.25860
0.50	0.42108
0.75	0.99588
1.00	1.06989
1.25	1.31500
1.50	1.46814
1.75	1.84658
2.00	2.24450

También nótese que el factor β aplica una escala sobre los campos completos, incluidas las regiones de activación, lo que quiere decir que en el funcional (2.14) se encontraría presente en el primer y tercer término, pero conviene dividir por β la expresión completa, de modo que solamente afecte al segundo término, donde se encuentra el parámetro λ , y dado que el valor de éste se obtiene por calibración, el valor β ya no se usa.

2.5.2. Calibración del nivel de activación a_1

El valor a_1 representa el nivel de activación en el modelo planteado en la ecuación (2.14), debido a esto es también el parámetro que permite regular la cantidad de falsos positivos obtenidos. En relación a esto último, se ha notado que la mayoría de los falsos positivos y falsos negativos obtenidos se encuentran ubicados en sitios

adyacentes a los bordes (o frontera) de las regiones de activación.

Considérese el experimento siguiente para demostrar la afirmación anterior. Sea la región de activación A definida como:

$$A = \{u \in \mathcal{L} : (u_2 - 24)^2 + (u_1 - 24)^2 \leq 64\}, \quad (2.41)$$

ubicada en una rejilla regular \mathcal{L} de dimensiones 50×50 , y el siguiente modelo de observación para el campo \hat{T} :

$$T(u) = n(u) + aI_A(u) \quad (2.42)$$

donde $n(u)$, $u \in \mathcal{L}$ son un conjunto de variables aleatorias independientes con distribución Gaussiana de media cero y varianza la unidad y I_A es la función indicadora de la región A .

Para 1000 corridas independientes, se utilizó un nivel de activación $a = 2$ y como parámetros $a_1 = 2$, $\lambda = 4$, $\nu = 0$. Al calcular el promedio de Verdaderos Positivos, Falsos Positivos y Falsos Negativos sobre cada elemento $u \in \mathcal{L}$, se obtuvieron los resultados mostrados en la Figura 2.6, donde se puede observar que los Falsos Positivos y Falsos Negativos resultantes se encuentran precisamente en la frontera de la región de activación.

De modo que si los errores de clasificación se encuentran principalmente sobre el borde de las regiones de activación, uno podría enfocarse en controlar los falsos positivos fuera de la frontera, lo cual se hace utilizando un nivel de activación $a = 0$ (sin agregar regiones de activación) y buscando el valor para el parámetro a_1 (nivel de activación del modelo) tal que cumpla la expresión siguiente:

$$a_1^* = \arg \min_{a_1} FPR(a, a_1) \text{ sujeto a } FPR(a, a_1) \leq \epsilon_1 \quad (2.43)$$

Donde ϵ_1 es un número positivo y pequeño. O equivalentemente, a_1^* es el valor que hace $FPR(0, a_1^*) = \epsilon_1$ ya que $FPR(0, a_1)$ es una función decreciente de a_1 .

Sin embargo, si es de interés realizar el control del FPR total, entonces se busca el valor a_1 tal que satisfaga:

$$FPR(a^*(a_1^*), a_1^*) = \epsilon_T \quad (2.44)$$

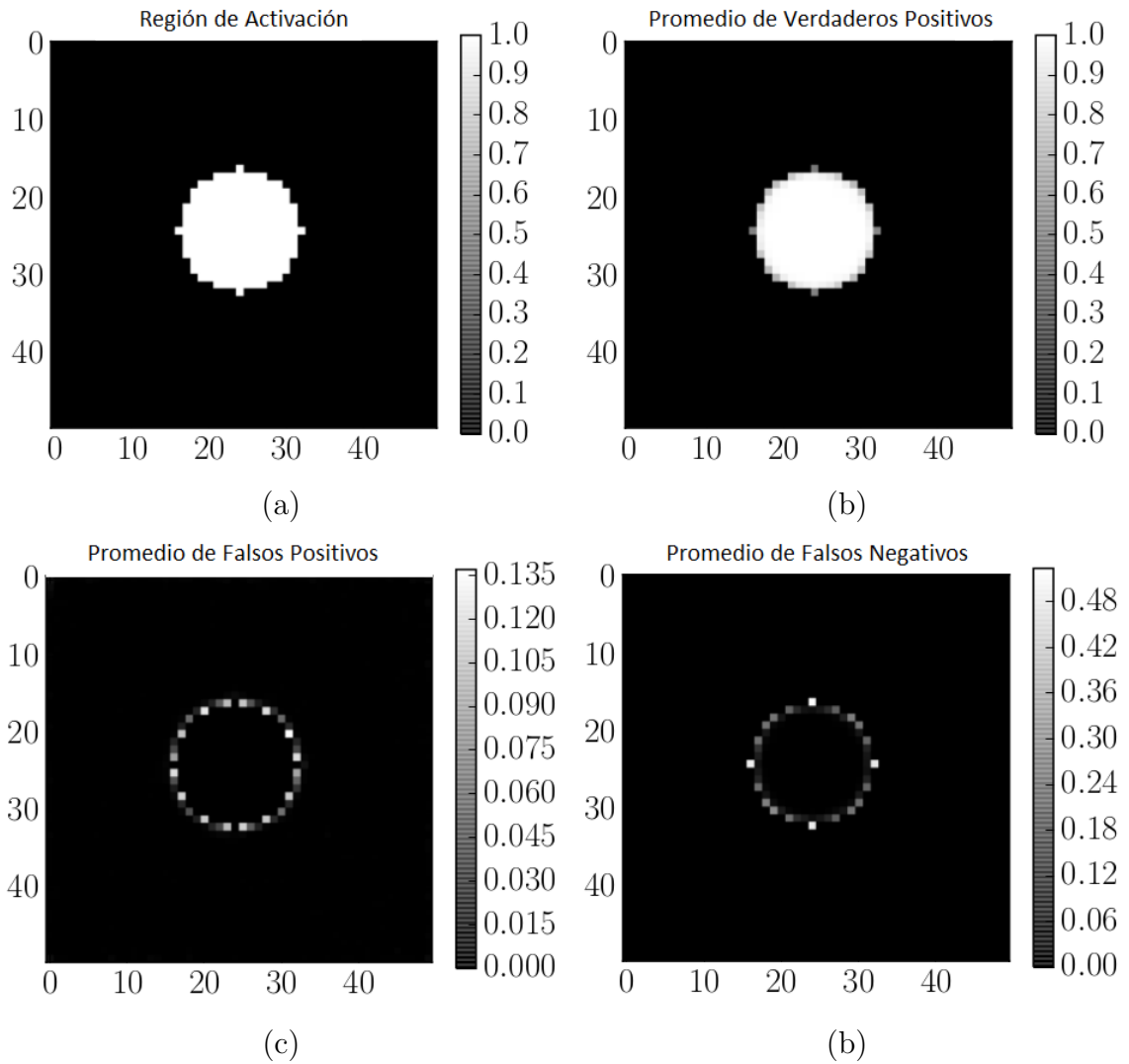


Figura 2.6: De izquierda a derecha, de arriba a abajo, región de activación, Promedio de Verdaderos Positivos, Falsos Positivos y Falsos Negativos

donde:

$$a^*(a_1) = \arg \max_a FPR(a, a_1) \quad (2.45)$$

Es decir, se busca un valor de a que sea el que maximice el FPR , lo equivalente a buscar el peor escenario, y encontrar el valor mínimo del parámetro a_1 tal que

controle el FPR a nivel ϵ_T .

El algoritmo 4 propone controlar inicialmente el FPR fuera de la frontera, es decir donde se busca simplemente encontrar el mínimo a_1 capaz de controlar el FPR a un nivel ϵ_1 sin distinguir entre los falsos positivos encontrados en la frontera de las regiones de activación o fuera de ella, como lo indica la ecuación (2.43). En una segunda etapa, a partir del valor a_1 encontrado, se define la región de activación (por ejemplo, la definida en (2.39)) y se va incrementando el nivel de activación a mientras el FPR se maximice, lo anterior expresado en la ecuación (2.45), luego con dicho nivel de activación fijo, se busca el mínimo valor a_1 que controle los falsos positivos dentro y fuera de la frontera, como indica la ecuación (2.44). Esta segunda etapa se realiza de manera iterativa, terminando cuando los valores a^* , a_1^* ya no se modifican. Nótese que en el procedimiento descrito, se tiene una cantidad $m < n$ de campos \hat{T} elegidos aleatoriamente, a los cuales se les agrega una región de activación sintética de nivel a y se segmenta a cada campo mediante la optimización de la funcional en (2.14), se calcula el FPR para cada campo y se usa el promedio de estos valores sobre los m campos.

Algoritmo 4 Calibración de a_1

Input: Un conjunto de n campos T generados bajo H_0 que han sido mapeados para tener una distribución normal estándar (\hat{T})

Output: a_1^* , el valor que cumple con las ecuaciones (2.43), (2.44) del parámetro a_1^* para el conjunto de campos \hat{T}

- 1: Inicializar el nivel de activación sintética en $a = 0$.
 - 2: Encontrar $a_1^* = \min a_1 : FPR(0, a_1^*) = \epsilon_1$ de la segmentación obtenida al resolver el funcional (2.14).
 - 3: Para el valor a_1^* obtenido en el paso anterior (es decir, fijar $a_1 = a_1^*$), encontrar $a^* = \arg \max FPR(a, a_1^*)$
 - 4: Para el valor a^* obtenido en el paso anterior (es decir, fijar $a = a^*$), encontrar $a_1^* = \min a_1 : FPR(a^*, a_1^*) = \epsilon_T$ de la segmentación obtenida al resolver el funcional (2.14).
 - 5: Repetir los pasos 3 y 4 hasta alcanzar convergencia.
-

2.5.3. Calibración de parámetro de regularización λ

En el modelo planteado en la ecuación (2.14) el parámetro λ representa el término de regularización para la correlación espacial de las regiones de activación, la elección de su valor no es un punto crítico en el desempeño del método siempre que este se encuentre dentro de un cierto intervalo, aunque éste sí depende del valor que toma el parámetro ν .

Para ilustrar la afirmación anterior, considérese el experimento siguiente. Uno puede generar campos T bajo H_0 con distinto grado de correlación utilizando Gibbs Sampler y el modelo en (2.12), variando el valor de ν en el modelo de ruido. Para cada uno de estos valores ν interesa encontrar λ^* , la λ tal que maximice la esperanza de verdaderos positivos, pero manteniendo control sobre los falsos positivos, como se expresa en la ecuación siguiente.

$$\lambda^*(\nu; \epsilon) = \arg \max_{\lambda} \overline{TPR}(\lambda; \epsilon) \quad (2.46)$$

donde:

$$\overline{TPR}(\lambda; \epsilon) = \frac{1}{N_a} \sum_{a_i \in a} TPR(a_i, \hat{a}_1, \lambda, \nu; \epsilon) \quad (2.47)$$

$$\approx \int_a TPR(a, \hat{a}_1, \lambda, \nu) P(a) da \quad (2.48)$$

Donde \hat{a}_1 satisface la ecuación (2.43). Es decir que, si uno prueba con diferentes valores para λ , en cada uno de ellos es necesario encontrar el valor de a_1 que cumpla con la ecuación (2.43), al que en este caso se nombra \hat{a}_1 . También nótese que la esperanza puede estimarse mediante el cálculo de la media estadística, a la que se nombra a partir de ahora \overline{TPR} , lo cual implica que se agregan regiones de activación sintéticas con niveles de activación a_i que toman valores en un conjunto C , $a_i \in C$, y de la segmentación resultante al optimizar el funcional (2.14) se calcula el TPR , se promedian estos valores usando (2.47), con N_a como la cardinalidad del conjunto de niveles de activación a para los que se calculó el TPR .

En la Figura 2.7 se muestra la λ^* encontrada para los valores de $\nu = \{0.0, 0.1, \dots, 3.5\}$, así como una función polinomial de grado 5 que ha sido ajustada a los datos a fin

de representar de una manera más general la relación entre los dos parámetros en cuestión, además de que permite interpolar entre valores no considerados en este experimento y obtener una aproximación de la λ^* correspondiente.

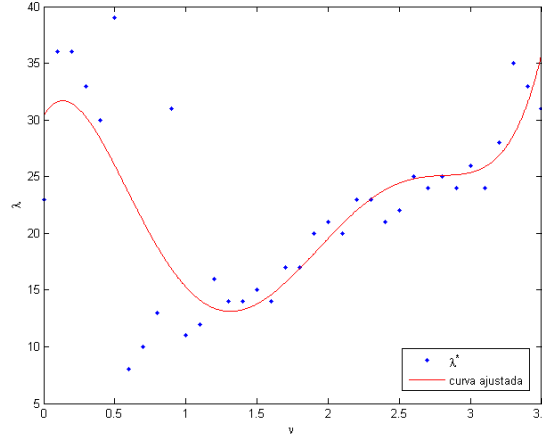


Figura 2.7: Curva ajustada $(3.02\nu^5 - 26.80\nu^4 + 82.25\nu^3 - 94.82\nu^2 + 21.24\nu + 30.36)$ para los datos ν vs. $\hat{\lambda}(\nu)$, para campos T generados mediante Gibbs Sampler usando ν modelo= $\{0.00, 0.25, \dots, 3.5\}$

A fin de analizar el desempeño del método al utilizar $\lambda = \lambda^*$ (donde λ^* es la λ óptima que resulta de la curva ajustada de la Figura 2.7) sobre campos con diferente correlación espacial, se planteó el experimento siguiente. Sea la región de activación A , definida en un espacio (o rejilla) regular \mathcal{L} de 50×50 :

$$A = u \in \mathcal{L} : (u_1 - 24)^2 + (u_2 - 24)^2 \leq 16, \quad (2.49)$$

un modelo de observación T bajo H_0 (T_0) planteado en la ecuación (2.12) utilizando Gibbs Sampler con ν modelo $\in C_\nu$, con $C_\nu = \{0, 0.25, \dots, 3.5\}$, al cual se le agrega una región de activación sintética usando:

$$T(u) = T_0(u) + aI_A(u) \quad (2.50)$$

con $u \in \mathcal{L}$, I_A es la función indicadora de la región A y a el nivel de activación impuesto, en este caso $a \in C_a$, con $C_a = \{0.0, 0.2, \dots, 5\}$. Para 1000 corridas indepen-

dientes y utilizando como parámetros a_1 tal que cumpla con la ecuación (2.43), es decir manteniendo bajo control el $FPR, \lambda = \lambda^*$, es decir la λ óptima que resulta de la curva del procedimiento de calibración (ver Figura 2.7) y ν considerando dos escenarios, el primero con $\nu = 0$ y el segundo usando ν es igual a la ν modelo (es decir, la misma utilizada en el modelo para generar los campos), descrito en la sección 2.5. Se calculó el promedio de TPR y se obtuvieron los resultados mostrados en la Figura 2.8, los cuales muestran que para campos cuyo valor de la ν modelo se encuentra entre cero y 0.75, el TPR reportado para ambos casos es muy similar. Sin embargo, conforme incrementa ν modelo, los resultados utilizando el parámetro $\nu = 0$ se ven afectados en mayor medida, el TPR promedio decrece más rápidamente, mientras que para los resultados usando ν igual a la ν modelo se observa una caída más lenta.

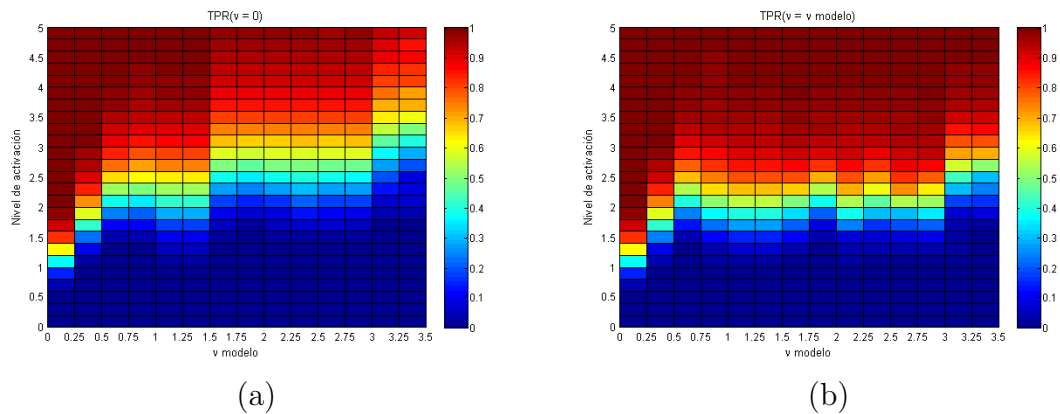


Figura 2.8: TPR calculado para campos T generados usando GS y el modelo en (2.12), con ν modelo $\in C_\nu$ y $C_\nu = \{0, 0.25, \dots, 3.5\}$, usando nivel de activación $a \in C_a$, con $C_a = \{0.0, 0.2, \dots, 5\}$. (a) $\nu = 0$, (b) $\nu = \nu^*$.

Uno puede analizar más fácilmente la ganancia obtenida cuando se utiliza el parámetro ν calculando la diferencia $TPR(\nu \text{ modelo}) - TPR(\nu = 0)$, como se muestra en la Figura 2.9, donde se observa que a partir de ν modelo igual a 0.75 se comienza a tener una pequeña ganancia en del $TPR(\nu \text{ modelo})$ sobre $TPR(\nu = 0)$, pero en realidad una ventaja significativa se observa a partir de ν modelo igual a 1.75.

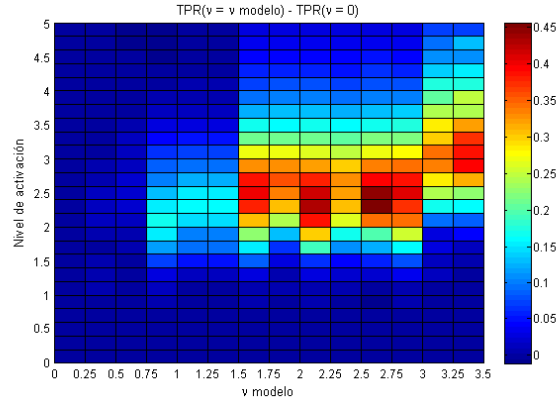


Figura 2.9: Diferencia $TPR(\nu = \nu \text{ modelo}) - TPR(\nu = 0)$, calculada para campos T generados usando GS y el modelo en (2.12), con $\nu \text{ modelo} \in C_\nu$ y $C_\nu = \{0, 0.25, \dots, 3.5\}$, usando nivel de activación $a \in C_a$, con $C_a = \{0.0, 0.2, \dots, 5\}$

2.6. Recapitulación

En esta sección se hace una síntesis del procedimiento propuesto para la detección de regiones de activación utilizando pruebas de hipótesis regularizadas.

Dado un valor pequeño ϵ , que representa la tolerancia que estamos dispuestos a aceptar para FPR , un modelo M_0 para generar muestras del estadígrafo T bajo H_0 (T_0) y un conjunto de campos observados $\{T^{(1)}, T^{(2)}, \dots, T^{(m)}\}$, la primera parte del procedimiento consiste en generar un conjunto de n campos T_0 , a partir de los cuales se obtiene la distribución nula. Ésta es utilizada en el paso 2 del procedimiento, que consiste en mapear cada campo T_0 para que los valores sigan una distribución normal estándar, de este modo se simplifica el cálculo de los campos de verosimilitud y los procedimientos de calibración de parámetros se vuelven universales, pues cualquiera que sea la distribución del estadígrafo que se esté utilizando, ésta se mapea siempre a una distribución conocida. Los siguientes tres pasos consisten en encontrar los parámetros adecuados para resolver la función de energía (2.14), estimar la ν^* utilizando el algoritmo 3, encontrar la λ^* de acuerdo con la curva ajustada en la sección 2.5.3, y encontrar la a_1 que cumpla con la condición (2.43) utilizando el algoritmo 4. Finalmente, se minimiza la función de energía 2.14 utilizando los parámetros encontrados y se declara como parte de la región de activación estimada \hat{A} a aquellos elementos u donde $b_1(u)$ es mayor a 0.5. El algoritmo 5 muestra de manera general

el procedimiento completo a seguir dados un conjunto de campos observados.

Algoritmo 5 Procedimiento completo

Input: ϵ , constante para controlar FPR . T , campos observados. M_0 , modelo para generar muestras del estadígrafo T bajo H_0 .

Output: \hat{A} , estimación de la región de activación.

- 1: Generar n campos del estadígrafo T bajo H_0 (T_0) usando el modelo M_0
- 2: Mapear los campos T_0 para que sus valores sigan una distribución normal estándar (\hat{T}_0), utilizando el algoritmo 1.
- 3: Estimar la ν óptima (ν^*) de los campos \hat{T}_0 utilizando el algoritmo 3.
- 4: Encontrar la λ óptima ($\lambda^*(\nu)$) de acuerdo con el procedimiento descrito en la subsección 2.5.3.
- 5: Seleccionar a_1^* utilizando el algoritmo de calibración 4.
- 6: Optimizar la función de energía (2.14) utilizando los parámetros ν^* , λ^* , a_1^* .
- 7: La región de activación estimada \hat{A} está compuesta por:

$$\hat{A} = \{u : b_1(u) > 0.5\} \tag{2.51}$$

Capítulo 3

Experimentos y resultados

En este capítulo se describen un conjunto de pruebas que se realizaron con el propósito de analizar el desempeño del método RHT, descrito en el capítulo anterior. Los experimentos se llevaron a cabo utilizando tanto datos sintéticos como datos fMRI, algunos de ellos enfocados a evidenciar el efecto de los parámetros involucrados en el método sobre el resultado final, además algunos de los resultados obtenidos fueron comparados para algunos casos con el método puntual FWER y con el método del estado del arte MBHT.

3.1. Pruebas con la selección parámetros

3.1.1. Calibración de λ

En esta subsección se presentan experimentos que muestran el impacto que tiene el parámetro de regularización λ sobre el TPR en datos sintéticos. Considérense un conjunto de campos T bajo H_0 (T_0) generados utilizando el modelo descrito en (2.12), mediante Gibbs Sampler usando ν modelo para los valores $\nu \in \{0.25, 0.5, 0.75, 1\}$. Luego, se utilizó una región de activación A , definida en (2.49), y se agregó a los campos mediante el siguiente modelo:

$$T(u) = T_0(u) + aI_A(u) \tag{3.1}$$

con $u \in \mathcal{L}$, donde \mathcal{L} es un espacio rectangular de 50×50 , I_A es la función indicadora de la región A y a el nivel de activación, que se probó para $a \in C_a$ con $C_a = \{0.0, 0.25, \dots, 5\}$. Para cada caso, se generaron 1000 campos independientes entre sí y se realizó la estimación de la región de activación mediante el método RHT con los parámetros ν igual al ν modelo utilizado para generar el campo y λ usando distintos valores, $\lambda = x/\sigma^2(\nu)$, $x \in C_x$, $C_x = \{2, 5, 10, 15, 20\}$. La Figura 3.1 muestra los resultados obtenidos, adicionalmente se incluye en la primera columna (indicada con *) los valores de TPR cuando se usan los parámetros $\nu = 0$ y $\lambda = 10$, se puede ver que en cada uno de los cuatro casos, la λ óptima corresponde a un valor diferente, que depende del valor ν modelo utilizado para generar los campos.

La Figura 3.2 muestra el TPR calculado para campos T bajo H_0 generados con el modelo (2.12) usando Gibbs Sampler con ν modelo = $\{0.0, 0.1, \dots, 3.5\}$, a los cuales se les agregaron regiones de activación sintéticas como se describe en la expresión (3.1), considerando tres casos para el nivel de activación $a \in C_a$, con $C_a = \{2, 2.5, 3.0\}$. Nuevamente, para 1000 campos independientes entre sí, se realizó la estimación de la región de activación mediante el método RHT con los parámetros ν igual al ν modelo utilizado para generar los campos y $\lambda = \{0, 1, \dots, 40\}$. En todos los casos se puede observar un comportamiento similar, al principio el incremento en el parámetro λ produce cambios significativos en el TPR , no obstante existe un rango de valores para los cuales se mantiene relativamente estable, para luego comenzar a decrecer.

La Figura 3.2 da cierta intuición acerca de los resultados al medir el TPR controlando FPR (con $\epsilon = 0.0001$), para campos con distinto grado de correlación en el ruido, se esperaría encontrar un intervalo de valores para λ en los cuales el método RHT obtiene resultados favorables. Uno se puede preguntar cómo afecta el valor ϵ sobre dichos resultados, en el siguiente experimento se considera el caso de calibración de λ con el parámetro $\nu = 0$, variando el valor $\epsilon \in C_\epsilon$, con $C_\epsilon = \{0.001, 0.0001, 0.00001\}$ y el valor del parámetro a_1 obtenido mediante el procedimiento de calibración (diferente para cada valor ϵ , ver subsección 2.5.2), utilizando campos cuya correlación en el ruido es resultado de aplicar un filtro gaussiano a campos de ruido gaussiano blanco, como se especifica en la expresión (2.40) usando un nivel de activación $a = 2$, con la región de activación A definida en la ecuación (2.49), en un espacio rectangular \mathcal{L} de 50×50 . Para 1000 campos generados independientes entre sí para una correla-

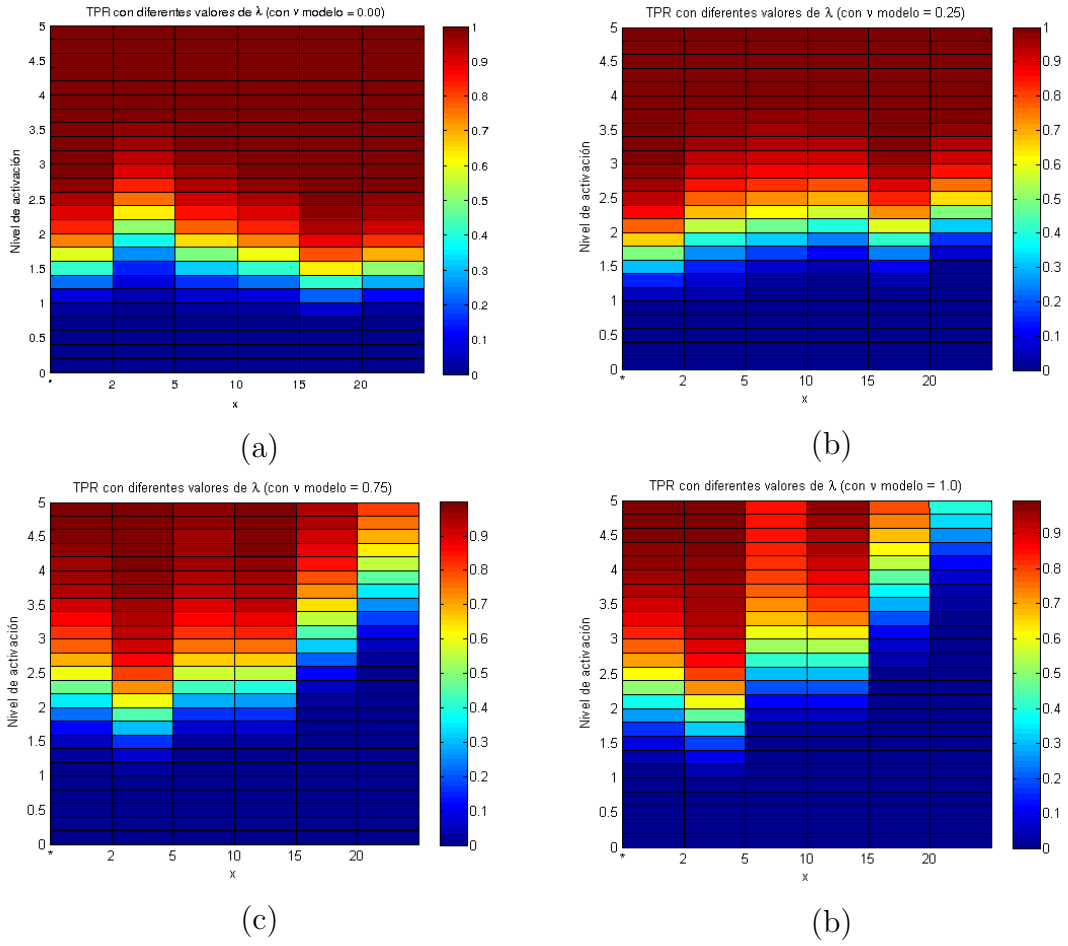


Figura 3.1: TPR obtenido con el método RHT para campos generados con el modelo (3.1), con ν modelo igual a a) 0, b) 0.25, c) 0.75, d) 1. La columna * corresponde a los resultados usando los parámetros $\nu = 0$ y $\lambda = 10$, el resto usando ν igual a la ν modelo con la que se generaron los campos, $\lambda = x/\sigma^2(\nu)$, $x \in C_x$, $C_x = \{2, 5, 10, 15, 20\}$, a_1 obtenida por el procedimiento de calibración descrito en la subsección 2.5.2 controlando $FPR \leq 0.0001$

ción de ruido $\rho = \{0.00, 0.05, 0.10, 0.15\}$ (la correlación se estimó de acuerdo con los datos mostrados en la tabla 2.1), se utilizó el método RHT (con los parámetros ya mencionados) para detectar la región de activación y se calculó el promedio de TPR obtenido con $\lambda \in \{0, 2, \dots, 32\}$. Los resultados aparecen en la Figura 3.3, las curvas obtenidas reflejan que existe un nivel de estabilidad al usar λ con valores entre 10 y 30 para los tres casos de ϵ probados, lo cual indica que la λ óptima se mantiene al

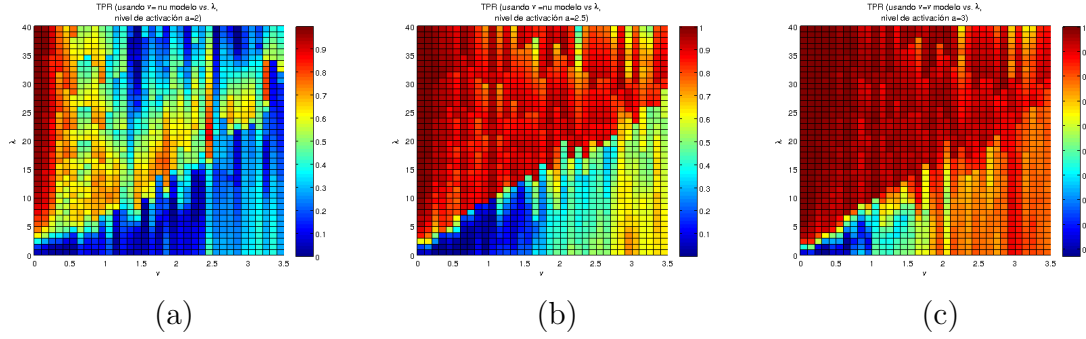


Figura 3.2: TPR obtenido con el método RHT para campos generados con el modelo (3.1) mediante GS, con ν modelo igual a $\{0.0, 0.1, \dots, 3.5\}$, usando como parámetros $\nu = \nu$ modelo , $\lambda = \{0, 1, \dots, 40\}$, a_1 obtenida por el procedimiento de calibración descrito en la subsección 2.5.2 controlando $FPR \leq 0.0001$, las regiones de activación sintética agregadas con niveles de activación: a)2, b) 2.5, c) 3.

variar el valor de ϵ .

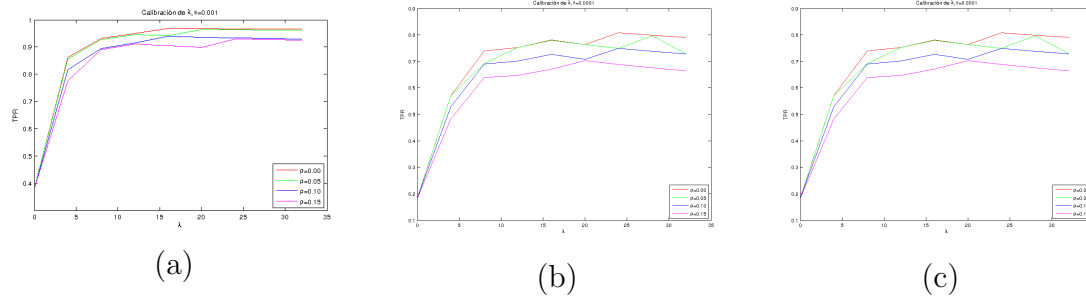


Figura 3.3: Curvas de TPR obtenidas al utilizar el método RHT para estimar región de activación de campos, usando los parámetros $\nu = 0$, $\lambda \in \{0, 2, \dots, 32\}$ y a_1 obtenida por el procedimiento de calibración, para campos generados usando el modelo (2.40) con diferente grado de correlación de ruido $\rho = \{0.00, 0.05, 0.10, 0.15\}$ (ver la tabla 2.1)

3.2. Comparación de desempeño

En el siguiente experimento, se tiene el propósito de comparar el desempeño del método propuesto (con y sin considerar el parámetro ν) contra el método puntual

FWER y el método del estado del arte MBHT. Para ello, se utilizaron campos T bajo H_0 (T_0) generados utilizando el modelo descrito en (2.12), mediante Gibbs Sampler con ν modelo igual a $\nu \in \{0, 0.5, 1, 1.5, 2\}$, a los cuales les fue agregada una región de activación A , definida en (2.49), usando:

$$T(u) = T_0(u) + aI_A(u) \quad (3.2)$$

con $u \in \mathcal{L}$, \mathcal{L} un espacio rectangular de dimensiones 50×50 , I_A es la función indicadora de la región A y a el nivel de activación impuesto, en este caso $a \in C_a$, con $C_a = \{0.0, 0.25, \dots, 5\}$.

Para el caso del método propuesto (RHT, pruebas de hipótesis regularizadas), se utilizaron como parámetros a_1 tal que cumpla con la ecuación (2.43), es decir manteniendo bajo control el FPR (se usó $\epsilon = 0.0001$), λ fija en 10, la primera con $\nu = 0$ y la segunda usando ν igual a la $\nu = \nu^*$, donde ν^* es la estimada utilizando el algoritmo 3, descrito en la sección 2.5. La Figura 3.4 muestra los resultados obtenidos, para campos generados usando ν modelo menor a 1 se puede ver un comportamiento muy similar entre las dos variantes de RHT evaluadas, ambos casos tienen resultados mejores a los del método puntual FWER en todos los niveles de activación probados en el experimento, también son superiores al método MBHT en la mayoría de los casos. Por otro lado, para campos generados usando ν modelo mayor o igual a 1 se observa mejora significativa de las dos variantes de RHT con respecto al desempeño del método puntual FWER, mientras que el método MBHT es ligeramente mejor para niveles de activación entre 0 y 1.5, pero en el resto nuevamente las dos variantes de RHT lo superan. En el caso de éstas últimas, la versión utilizando $\nu = \nu^*$ obtiene resultados más favorables para niveles de activación entre 2.5 y 4, cabe mencionar que en este caso se ha dejado el hiperparámetro λ fijo en un mismo valor para ambos casos.

Nótese también que hasta este punto los niveles de activación utilizados se encuentran en el rango de valores esperados, uno pudiera preguntarse el efecto que tendría en el desempeño del método si se utilizaran niveles de activación extremos (muy bajos o muy altos), a tal grado que el valor de la verosimilitud $l_k(u)$ para las clases activación y no activación fuera muy cercano. Lo anterior pudiera representar un problema en la clasificación, además de que el peso del término de regulariza-

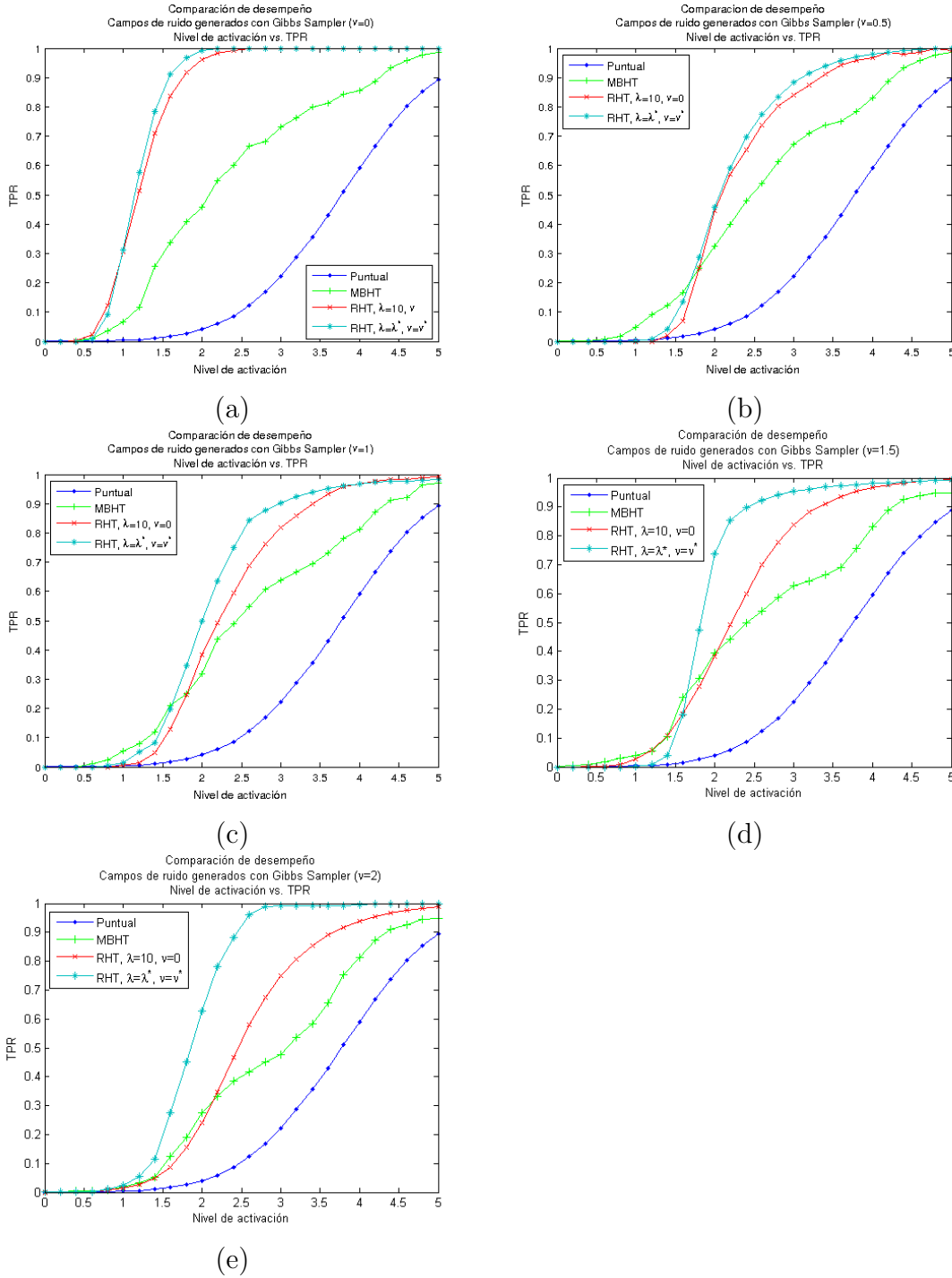


Figura 3.4: Comparación de desempeño (nivel de activación vs. TPR) para campos generados utilizando el modelo en (2.12), usando GS. (a) usando ν modelo = 0, (b) usando ν modelo = 0.5, (c) usando ν modelo = 1, (d) usando ν modelo = 1.5, (e) usando ν modelo = 2 .

ción λ perdería importancia pues $-\log l_k(u)$ sería mucho mayor. Una manera de dar solución a esta dificultad puede ser utilizar una normalización sobre el término de verosimilitud, el siguiente experimento tiene el objetivo de observar el desempeño de RHT para niveles de activación más altos y comparar los resultados para dos variantes, la primera usando la verosimilitud sin normalizar y la segunda usando la verosimilitud normalizada, en ambos casos usando como parámetros $\nu = 0, \lambda = \lambda^*$ y a_1 (se usó $\epsilon = 0.0001$) escogida usando el procedimiento de calibración. Nuevamente, se utilizaron campos T bajo $H_0 (T_0)$ generados utilizando el modelo descrito en (2.12) mediante Gibbs Sampler, con ν modelo igual a $\nu \in \{0, 1.5\}$, a los cuales les fue agregada una región de activación A , definida en (2.49), y usando el modelo de observación (3.2), con $u \in \mathcal{L}$, \mathcal{L} un espacio rectangular de dimensiones 50×50 , I_A es la función indicadora de la región A y a el nivel de activación impuesto, en este caso $a \in C_a$, con $C_a = \{0.0, 0.25, \dots, 10\}$. La Figura 3.5 muestra los resultados promedio obtenidos para 1000 campos generados independientes entre sí, en ambas gráficas se observa que la clasificación no se ve afectada al incrementar el nivel de activación en las regiones y tampoco existe mejoría al usar la normalización en la verosimilitud.

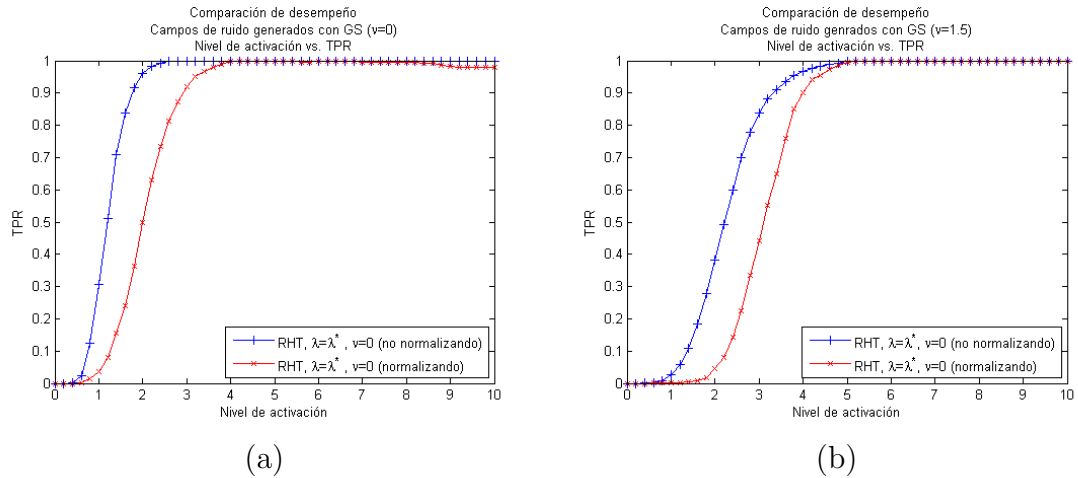


Figura 3.5: Comparación de desempeño (nivel de activación vs. TPR) para campos generados utilizando el modelo en (2.12), usando GS. (a) usando ν modelo = 0, (b) usando ν modelo = 1.5, en azul usando la verosimilitud no normalizada, en rojo la verosimilitud normalizada

3.3. Pruebas con datos fMRI

En esta sección se presentan experimentos utilizando un conjunto de datos fMRI liberados por el University College de Londres (UCL) para ser utilizados en fines educativos. En las subsecciones siguientes se describen a detalle las características de dichos datos, se analiza la distribución de ν para el caso específico de datos fMRI y se observa el desempeño del método RHT al variar la relación señal ruido.

3.3.1. Datos fMRI de activación auditiva

Los experimentos descritos en esta parte del documento fueron realizados utilizando un conjunto de datos recolectados por el Laboratorio de Imagenología Funcional (FIL) del UCL, bajo la supervisión de Karl Friston, los cuales se encuentran disponibles para propósitos educativos en el sitio web del FIL (ver Rees et al. (2015)). El experimento se realizó a un solo sujeto, los datos corresponden a 96 volúmenes (serie fM00223), compuestos de 64 rebanadas contiguas (64x64x64 voxeles de 3mm x 3mm x 3mm), que fueron adquiridos en bloques de 6 y donde el tiempo de repetición (RT) entre escaneo se estableció en 7 segundos. Es decir, en total 16 bloques de 42 s, aunque debido a efectos relativos al T1 se ha seguido la recomendación de descartar los primeros dos bloques. La condición para bloques sucesivos alternó entre descanso y estimulación auditiva, comenzando con descanso. La estimulación auditiva consistió de palabras de dos sílabas presentadas binauralmente a una frecuencia de 60 por minuto. Además de lo anterior, los datos incluyen una imagen anatómica, que se utiliza en el documento para la visualización de resultados.

3.3.2. Estimación de la distribución de ν

Es de interés conocer la distribución la ν para datos de éste tipo, en este sentido, un primer experimento consistió en tomar datos de fMRI (de SPM) y realizar la estimación de ν para cada rebanada de volumen. Considérese que en este caso la región \mathcal{L} no es un espacio rectangular, y tampoco es la misma para cada rebanada, sino que adopta la forma de la materia gris del sujeto observado. Esto quiere decir que existen rebanadas en las cuales la cantidad de elementos u considerados es menor, por ejemplo las rebanadas de los extremos del volumen contienen pocos elementos,

mientras que las rebanadas más próximas al centro contienen más. De modo que para obtener la distribución, la estimación de ν resultante de cada rebanada fue ponderada por el número de elementos en la región \mathcal{L} . Luego, la distribución obtenida se normaliza para obtener la función de probabilidad, que se muestra en la Figura 3.6, donde se distingue que los valores con mayor probabilidad de ocurrencia se encuentran entre 0 y 1.

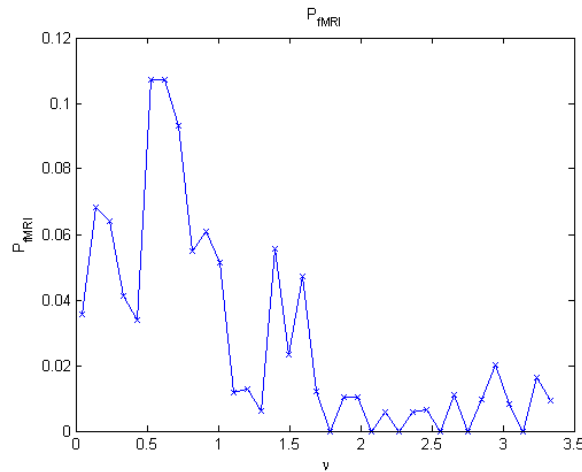


Figura 3.6: Función de probabilidad de ν para datos fMRI

Por otro lado, retomando los resultados obtenidos en el experimento de la sección 2.5.3, donde se calcula el TPR para campos generados usando el modelo 2.12 con GS y diferentes ν modelo y niveles de activación, y se analiza la ganancia obtenida al utilizar el método RHT con el parámetro ν en cero o usando ν modelo, uno puede nombrar $\Delta TPR = TPR(\nu = \nu \text{ modelo}) - TPR(\nu = 0)$ (mostrada en la Figura 2.9) y luego tomar el máximo por columnas, es decir la $\max(\Delta TPR)$ para cada ν , con lo que se obtiene la función que se muestra en la Figura 3.7, de la cual se puede notar que la mayor ganancia se encuentra en aproximadamente 0.45 y obtiene para ν entre 1.5 y 3, mientras que para valores menores a 1.5 la ganancia es más modesta, lo que sugiere que para datos fMRI el beneficio de usar ν no será tan significativo.

Para probar lo anterior, uno puede calcular la ganancia esperada $E[g]$ para datos

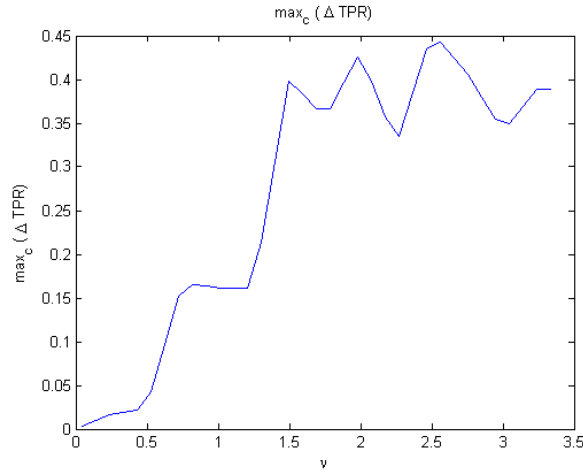


Figura 3.7: $\max_c(\Delta TPR)$

fMRI calculando:

$$E[g] = \int_{-\infty}^{\infty} \max_c \Delta TPR(\nu) P_{fMRI}(\nu) d\nu \quad (3.3)$$

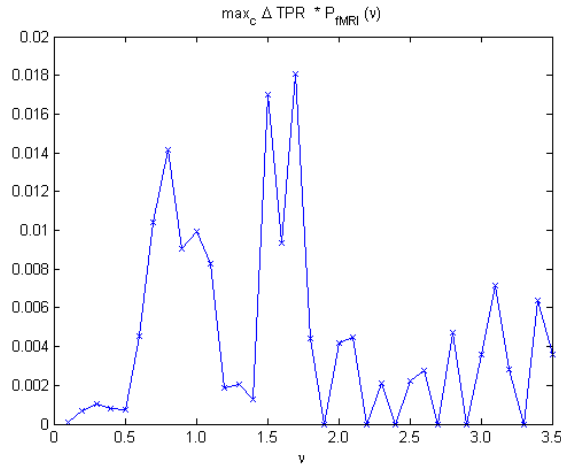


Figura 3.8: $\max_c \Delta TPR(\nu) P_{fMRI}(\nu)$

La curva $\max_c \Delta TPR(\nu) P_{fMRI}(\nu)$ se muestra en la Figura 3.8, la ganancia esperada resultante al calcular el área bajo ésta curva es de 0.0151, lo que indica que al utilizar la ν estimada se espera un incremento de 1.5% en el TPR. Esta ganancia

es pequeña, los experimentos revisados hasta este punto señalan que para los datos fMRI conviene simplificar el procedimiento utilizando $\nu = 0$, pues los resultados con este valor son muy similares a los obtenidos utilizando la estimación de ν obtenida usando el algoritmo 3.

3.3.3. Variación de la constante ϵ de control de FPR (o mapa de p -valores generalizados)

Una manera común de presentar resultados en imágenes de resonancia magnética funcional cuando se usan métodos puntuales es construir un mapa de significancia con colores, donde cada color represente las regiones de activación detectadas a un nivel α distinto, con α como la probabilidad de cometer uno o más errores de tipo I (falsos positivos) en las n pruebas, o bien graficar el p -valor sin umbralizar utilizando una escala de colores. La construcción e interpretación de este tipo de gráficas en el caso de los métodos puntuales es directa, pues para declarar a un elemento como activo se utiliza únicamente la magnitud del estadígrafo en esa posición, a diferencia otros enfoques (como el método RHT) que consideran la magnitud de los elementos vecinos para tomar una decisión. En el caso específico del método RHT, aunque no existe como tal el concepto de nivel de significancia, uno puede construir una gráfica equivalente considerando lo siguiente: si se toma en cuenta que en el caso de los métodos puntuales el valor α influye sobre la decisión de rechazar la hipótesis nula cuando es verdadera, es decir que controla el error tipo I (falsos positivos), uno puede decir que el valor ϵ (ver subsección 2.5.2) en el método RHT funciona de manera análoga, pues su valor influye directamente en el proceso de calibración del parámetro a_1 , pues se escoge tal que controle el FPR a un nivel ϵ .

De acuerdo con lo anterior, la prueba siguiente consiste en observar el comportamiento del método RHT al variar la constante ϵ que controla el FPR sobre los resultados. Para ello, utilizando el estadígrafo F , se tomaron los datos del experimento de activación auditiva (ver 3.3.1) y se realizó la estimación de regiones de activación con el método RHT usando los parámetros $\nu = 0$, $\lambda = 20$ y a_1 obtenida por el procedimiento de calibración (ver 2.5.2) tal que cumpla con la ecuación (2.43), para $\epsilon \in \{0.01, 0.001, 0.0001, 0.00001\}$. El mapa de *significancia*, o mapa de *p -valores generalizados*, se obtiene al sobreponer las regiones de activación detecta-

das utilizando distintos valores para ϵ , como se muestra en la Figura 3.9. La gráfica muestra diferencias importantes entre el primer y segundo nivel, pero en los últimos niveles las regiones detectadas son más similares entre sí, con diferencias menores solamente en la frontera.

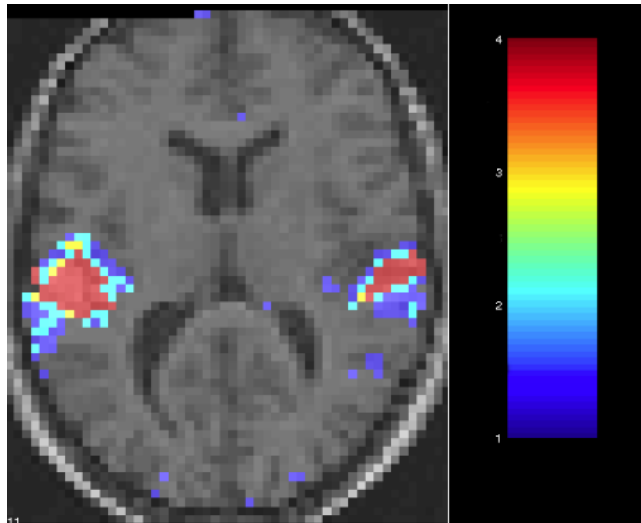


Figura 3.9: Mapa de p-valores generalizados obtenido al sobreponer las regiones de activación detectadas mediante el método RHT, usando los parámetros $\nu = 0$, $\lambda = 20$ y a_1 obtenida por el procedimiento de calibración (ver 2.5.2) controlando $FPR \leq \epsilon$, utilizando distintos valores para ϵ : 1) 0.01, 2) 0.001, 3) 0.0001, 4) 0.00001

3.3.4. Reducción del número de bloques

El siguiente experimento se realizó con el fin de probar la estabilidad del método RHT ante la variación de la relación señal-ruido, comparado con el comportamiento del método del estado del arte MBHT y del método puntual FWER, para lo cual se tomaron los datos del experimento de activación auditiva, compuestos originalmente de 14 bloques (los primeros 2 se descartan por recomendación, como ya se ha mencionado a detalle en el apartado 3.3.1). Se realizó la estimación de regiones de activación utilizando el estadígrafo F para diferente número de bloques, en cada caso reduciendo el número de bloques en dos, iniciando en 14 hasta llegar a 4. Para el método RHT, se utilizaron los parámetros $\nu = 0$, $\lambda = 20$ y a_1 obtenida por el procedimiento de calibración. Para los tres métodos (RHT, MBHT y puntual), se ha

controlado $FPR \leq 0.0001$. La Figura 3.10 muestra las regiones de activación detectadas utilizando 14, 10, 6 y 4 bloques, las comparaciones indican que en el método puntual se presenta mayor afectación de las regiones estimadas al reducir el número de bloques, al grado de que al utilizar solamente 6 bloques las regiones casi han desaparecido. Tanto MBHT como RHT obtienen mejores resultados y muestran mayor robustez a la reducción de información para la mayoría de los casos, sin embargo para el caso cuando se utilizan solamente cuatro bloques MBHT no detecta ninguna región de activación, el método puntual apenas detecta una región y el método RHT alcanza a recuperar dos regiones de activación, superando así a los otros dos.

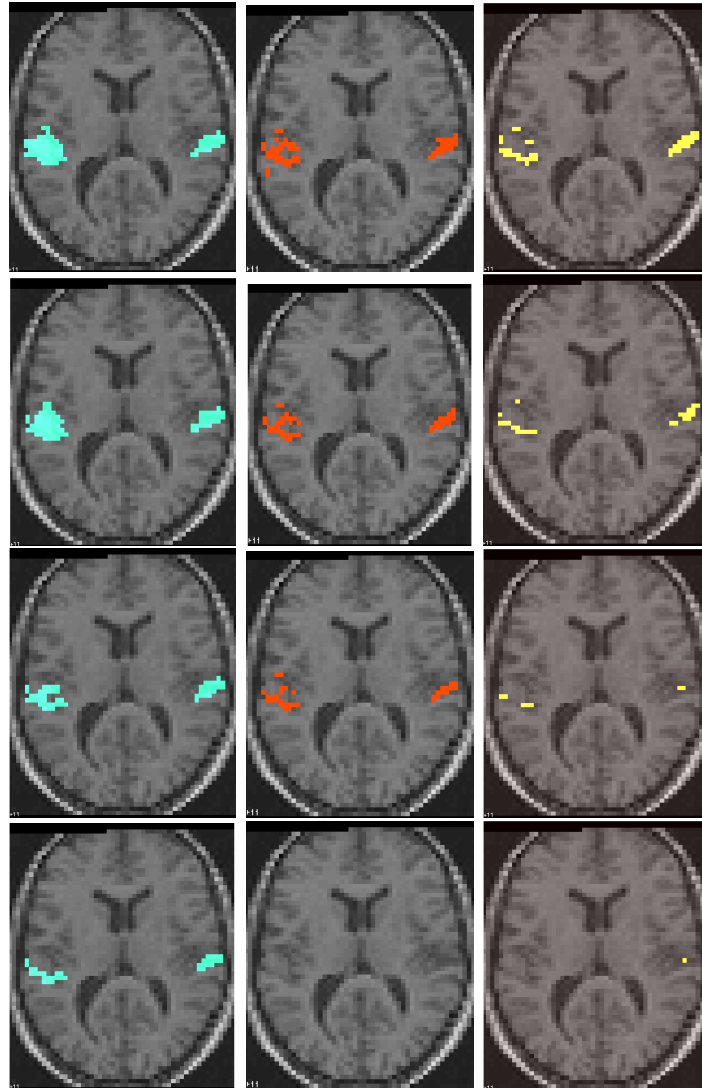


Figura 3.10: Comparación de regiones de activación detectadas, de arriba a abajo: 14, 10, 6 y 4 bloques. A la izquierda, regiones de activación estimadas con RHT usando el estadígrafo F (usando $\lambda = 20$ y a_1 obtenida por calibración controlado $FPR \leq 0.0001$). Al centro, regiones de activación estimadas usando el método del estado del arte MBHT usando el estadígrafo F. A la derecha, regiones de activación estimadas usando el método puntual FWER, también usando el estadígrafo F.

Capítulo 4

Conclusiones

En el trabajo presentado se expuso un nuevo método para la detección de regiones de activación en campos bidimensionales discretos que mejora notablemente el estado del arte, el procedimiento de pruebas de hipótesis regularizadas aborda el problema de la detección de regiones activas como un asunto de segmentación y está basado en el algoritmo de QMMF propuesto por (Rivera et al., 2007), modelando los campos del estadígrafo bajo H_0 (T_0) bajo un enfoque bayesiano y obteniendo una aproximación de la distribución marginal a posteriori de la función indicadora de la región de activación A , a continuación se hace mención de los puntos más importantes del método formulado en este documento.

El método RHT, presentado en el capítulo 2, fue desarrollado para dos casos, el caso sencillo considera que se tiene ruido blanco gaussiano aditivo, el segundo caso asume que se tiene ruido correlacionado espacialmente, para lo cual se presentó un nuevo enfoque que considera esta característica como parte del modelo mismo (ver 2.2.2), en este último escenario se agrega el parámetro ν a la función de energía a minimizar, el valor de ν está relacionado con el grado de correlación de los campos T_0 . Además de ν , otros parámetros se encuentran presentes: a_1 (nivel de activación en el modelo) y λ (término de regularización, controla a granularidad de las regiones de activación detectadas) y se propuso un método de calibración universal de dichos parámetros mediante un procedimiento de mapeo de la distribución del estadígrafo bajo H_0 a una distribución normal estándar (subsección 2.4), aprovechando algunas propiedades de la distribución gaussiana y la función error inversa, lo cual hace que

el proceso de selección de los parámetros se vuelva general, ya que se trabaja siempre con campos cuya distribución es normal estándar y entonces el método no depende del estadígrafo que se decida utilizar.

En lo que se refiere al término de correlación espacial del ruido, se mostró que el valor óptimo de ν se puede estimar a través del cálculo de su pseudoverosimilitud, sin embargo, al ser mapeados los campos a una distribución normal estándar la varianza original en los datos se modifica y dificulta la estimación directa de ν , pues existe una relación implícita entre estos dos valores, por lo que se propuso un procedimiento basado en el algoritmo Esperanza-Maximización (subsección 2.5.1), que iterativamente busca recuperar la consistencia entre la ν estimada y la varianza de los datos. Además, experimentalmente se mostró que para campos cuya ν estimada se encuentra entre cero y uno, la ganancia obtenida del método utilizando la ν óptima con respecto a utilizar $\nu = 0$ es muy pequeña, por lo que es preferible simplificar el método utilizando $\nu = 0$. En cambio, para datos con ν mayores a 1, la ganancia es notoria para niveles de activación entre 1.5 y 3.5.

En los experimentos presentados en el capítulo 3 se mostró que el valor óptimo (λ^*) del término de regularización, es decir el valor que maximiza el TPR manteniendo bajo control el FPR , se ve influenciado por el valor de la ν estimada, por lo que λ^* se debe elegir en función de ν , no obstante se mostró que el método RHT no es muy sensible al parámetro λ , pues existe un rango de valores (diferente para ν) para los cuales se obtienen resultados favorables, además de que en la practica uno puede calcular λ^* para varias ν y luego ajustar una curva como en la Figura 2.7. Además, el método propuesto fue comparado en desempeño (TPR vs. Nivel de activación) para campos correlacionados generados usando el modelo (2.27), la comparación se hizo contra los métodos con el método puntual FWER y con el método del estado del arte MBHT, los resultados obtenidos mostraron que comparado con el método puntual las dos variantes de RHT (una usando la ν^* estimada y la otra usando $\nu = 0$) fueron superiores en prácticamente todos los niveles de activación evaluados. Además, la comparación con respecto a MBHT indicó un desempeño superior por parte del método RHT para niveles de activación entre 1.6 y 4 utilizando la ν^* estimada, mientras que utilizando el parámetro $\nu = 0$ se obtuvieron mejoras visibles con niveles de activación entre 2.6 y 4.

Por otro lado, de los experimentos realizados con datos de fMRI (subsección 3.3.1), se realizó la estimación de la función de probabilidad del término de correlación espacial del ruido en imágenes de fMRI, a partir de la cual se calculó la ganancia máxima esperada de TPR al incluir el parámetro ν y se obtuvo que hay una mejora de 1.5% en el TPR , lo que indica que en el caso de datos fMRI conviene simplificar el procedimiento utilizando $\nu = 0$, sin embargo para datos cuya ν estimada sea mayor a 1.5 sí es apropiado utilizar dicha estimación. Además de lo anterior, se presentó una alternativa a los mapas de significancia construidos en los métodos puntuales, nombrada mapa de *p-valores generalizados*, que consiste en graficar las regiones de activación detectadas a diferentes grado de tolerancia del nivel de FPR sobre una misma imagen, de modo que se aprecie el efecto que tiene sobre las regiones el controlar a diferente medida el error tipo I. También se evaluó la estabilidad del método mediante una comparación de regiones de activación detectadas al variar la relación señal-ruido, usando el método puntual y el método propuesto, siendo éste último el que presentó mejor comportamiento al sufrir menos afectación en las regiones detectadas.

Como trabajo a futuro se espera realizar la implementación del método propuesto en una herramienta de software para el análisis de imágenes fMRI, tales como SPM o FSL, de modo que se pueda compartir la idea desarrollada en este documento y recibir realimentación del método. Asimismo, sería interesante obtener resultados para conjuntos de datos reales adicionales a los aquí presentados, a fin de verificar la robustez del método.

Apéndice A

Cálculo del gradiente para la funcional en la ec. (2.14)

Sean:

$$d_k(u) = (T(u) - a_k) \quad (\text{A.1})$$

$$p(u) = b_1(u) \quad (\text{A.2})$$

Retomando la ec. (2.14):

$$U(p) = \frac{1}{2} \sum_u \sum_k d_k^2(u) b_k^2(u) + \lambda \sum_{\langle u,v \rangle} (b_k(u) - b_k(v))^2 \quad (\text{A.3})$$

$$+ \nu \sum_{\langle u,v \rangle} \sum_{i,j} (d_i(u) - d_j(v))^2 b_k(u) b_j(v) \quad (\text{A.4})$$

$$\text{Sujeto a: } \sum_{k=0}^N b_k(u) = 1 \quad (\text{A.5})$$

$$b_k(u) \geq 0 \quad (\text{A.6})$$

Desarrollando la ec. anterior:

$$U(p) = \frac{1}{2} \sum_u [d_0^2(u)(1-p(u))^2 + d_1^2(u)p(u)] \quad (\text{A.7})$$

$$+ 2\nu \sum_{\langle u,v \rangle} [[d_0(u) - d_0(v)]^2(1-p(u))(1-p(v))] \quad (\text{A.8})$$

$$+ [d_0(u) - d_1(v)]^2(1-p(u))p(v) \quad (\text{A.9})$$

$$+ [d_1(u) - d_0(v)]^2p(u)(1-p(v)) \quad (\text{A.10})$$

$$+ [d_1(u) - d_1(v)]^2p(u)p(v) \quad (\text{A.11})$$

$$+ \lambda \sum_{\langle u,v \rangle} [[(1-p(u)) - (1-p(v))]^2] \quad (\text{A.12})$$

$$+ [p(u) - p(v)]^2 \quad (\text{A.13})$$

Luego,

$$\frac{\partial U(p)}{\partial p(u)} = d_0^2(u)p(u) + d_1^2(u)p(u) - d_0^2(u) \quad (\text{A.14})$$

$$+ \lambda \sum_{v \in N_u} 2([p(u) - p(v)] - [p(v) - p(u)]) \quad (\text{A.15})$$

$$+ \nu \sum_{v \in N_u} [-[d_0(u) - d_0(v)]^2(1-p(v))] \quad (\text{A.16})$$

$$- [d_0(u) - d_1(v)]^2p(v) + [d_1(u) - d_0(v)]^2(1-p(v)) \quad (\text{A.17})$$

$$+ [d_1(u) - d_1(v)]^2p(v) \quad (\text{A.18})$$

$$= d_0^2(u)p(u) + d_1^2(u)p(u) - d_0^2(u) \quad (\text{A.19})$$

$$+ 4\lambda \sum_{v \in N_u} [p(u) - p(v)] \quad (\text{A.20})$$

$$+ \nu \sum_{v \in N_u} [2a(u, v)p(v) + b(u, v)] \quad (\text{A.21})$$

$$= 0 \quad (\text{A.22})$$

Donde

$$a(u, v) = [d_0(u) - d_0(v)]^2 - [d_0(u) - d_1(v)]^2 \quad (\text{A.23})$$

$$- [d_1(u) - d_0(v)]^2 + [d_1(u) - d_1(v)]^2 \quad (\text{A.24})$$

$$b(u, v) = [d_1(u) - d_0(v)]^2 - [d_0(u) - d_0(v)]^2 \quad (\text{A.25})$$

A.0.5. Simplificaciones

Si desarrollamos los cuadrados de $a(u, v)$ (Ec. (6)):

$$a(u, v) = d_0^2(u) - 2d_0(u)d_0(v) + d_0^2(v) \quad (\text{A.26})$$

$$- d_0^2(u) + 2d_0(u)d_1(v) + d_1^2(v) \quad (\text{A.27})$$

$$- d_1^2(u) + 2d_1(u)d_0(v) + d_0^2(v) \quad (\text{A.28})$$

$$+ d_1^2(u) - 2d_1(u)d_1(v) + d_1^2(v) \quad (\text{A.29})$$

$$= 2[-d_0(u)d_0(v) + d_0(u)d_1(v)] \quad (\text{A.30})$$

$$+ d_1(u)d_0(v) - d_1(u)d_1(v)] \quad (\text{A.31})$$

$$a(u, v) = -d_0(u)d_0(v) + d_0(u)d_1(v) + d_1(u)d_0(v) - d_1(u)d_1(v) \quad (\text{A.32})$$

$$= d_0(u)[-d_0(v) + d_1(v)] + d_1(u)[d_0(v) - d_1(v)] \quad (\text{A.33})$$

$$= d_0(u)[d_1(v) - d_0(v)] - d_1(u)[d_1(v) - d_0(v)] \quad (\text{A.34})$$

$$= -[d_1(u) - d_0(u)][d_1(v) - d_0(v)] \quad (\text{A.35})$$

Por otro lado

$$d_1(u) - d_0(u) = (T(u) - a_1) - (T(u) - a_0) = a_0 - a_1 = -a_1 \quad (\text{A.36})$$

$$d_1(v) - d_0(v) = (T(v) - a_1) - (T(v) - a_0) = a_0 - a_1 = -a_1 \quad (\text{A.37})$$

Por tanto

$$a(u, v) = -a_1^2 \quad (\text{A.38})$$

$$b(u, v) = [d_1(u) - d_0(v)]^2 - [d_0(u) - d_0(v)]^2 \quad (\text{A.39})$$

$$= [d_1(u) - d_0(v) + d_0(u) - d_0(v)] \quad (\text{A.40})$$

$$[d_1(u) - d_0(v) - d_0(u) + d_0(v)] \quad (\text{A.41})$$

$$= [(T(u) - a_1) - (T(v) - a_0)] \quad (\text{A.42})$$

$$+(T(u) - a_0) - (T(v) - a_0)][d_1(u) - d_0(u)] \quad (\text{A.43})$$

$$= -a_1[(T(u) - a_1) - (T(v) - a_0)] \quad (\text{A.44})$$

$$+(T(u) - a_0) - (T(v) - a_0)] \quad (\text{A.45})$$

$$= -a_1[2T(u) - 2T(v) - a_1] \quad (\text{A.46})$$

$$= -2a_1[T(u) - T(v)] + a_1^2 \quad (\text{A.47})$$

$$= -2a_1\Delta T(u, v) + a_1^2 \quad (\text{A.48})$$

con $\Delta T(u, v) = T(u) - T(v)$.

Apéndice B

Pseudoverosimilitud de ν

El modelo de ruido correlacionado es el siguiente:

$$P_n(n) = \frac{1}{Z} \exp(-U_n(n)) \quad (\text{B.1})$$

con:

$$U_n(n) = \frac{1}{2} \sum_u n^2(u) + \nu \sum_{\langle u,v \rangle} (n(u) - n(v))^2 \quad (\text{B.2})$$

Luego, para estimar el parámetro ν , se calcula la máxima pseudoverosimilitud del siguiente modo:

$$P(n(u)|n(v), v \in N_u) = \frac{\exp(-U_u(n(u)))}{\int_{-\infty}^{\infty} \exp(-U_u(x)) dx} \quad (\text{B.3})$$

donde:

$$U_u(x) = \frac{1}{2}x^2 + \nu \sum_{v \in N_u} (x^2 - 2xn(v) + n^2(v)) \quad (\text{B.4})$$

Nótese que la integral en el denominador de la ecuación (4) es similar a la función de densidad acumulativa de la distribución normal:

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1 \quad (\text{B.5})$$

$$\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \sigma\sqrt{2\pi} \quad (\text{B.6})$$

Entonces, completando el trinomio cuadrado perfecto en la expresión (B.4):

$$U_u(x) = \frac{1}{2}x^2 + \nu|N_u|x^2 - 2x\nu \sum_{v \in N_u} n(v) + \nu \sum_{v \in N_u} n^2(v) \quad (\text{B.7})$$

$$= \left(\frac{1}{2} + \nu|N_u|\right)x^2 - 2x\nu \sum_{v \in N_u} n(v) + \nu \sum_{v \in N_u} n^2(v) \quad (\text{B.8})$$

$$= \left(\frac{1}{2} + \nu|N_u|\right) \left(x^2 - \frac{2x\nu \sum_{v \in N_u} n(v)}{\frac{1}{2} + \nu|N_u|} \right) + \nu \sum_{v \in N_u} n^2(v) \quad (\text{B.9})$$

$$= \left(\frac{1}{2} + \nu|N_u|\right) \left(x^2 - \frac{2x\nu \sum_{v \in N_u} n(v)}{\frac{1}{2} + \nu|N_u|} + \frac{\left(\nu \sum_{v \in N_u} n(v)\right)^2}{\left(\frac{1}{2} + \nu|N_u|\right)^2} \right) \quad (\text{B.10})$$

$$+ \nu \sum_{v \in N_u} n^2(v) - \frac{\left(\nu \sum_{v \in N_u} n(v)\right)^2}{\frac{1}{2} + \nu|N_u|} \quad (\text{B.11})$$

$$= \left(\frac{1}{2} + \nu|N_u|\right) \left(x - \frac{\nu \sum_{v \in N_u} n(v)}{\frac{1}{2} + \nu|N_u|} \right)^2 \quad (\text{B.12})$$

$$+ \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2 \left(\sum_{v \in N_u} n(v)\right)^2}{\frac{1}{2} + \nu|N_u|} \quad (\text{B.13})$$

$$= \frac{(x - \mu_x)^2}{2\sigma_x^2} + B_x \quad (\text{B.14})$$

con:

$$\mu_x = \frac{\nu \sum_{v \in N_u} n(v)}{\frac{1}{2} + \nu |N_u|} \quad (\text{B.15})$$

$$\sigma_x = \frac{1}{\sqrt{2(\frac{1}{2} + \nu |N_u|)}} \quad (\text{B.16})$$

$$B_x = \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2 \left(\sum_{v \in N_u} n(v) \right)^2}{\frac{1}{2} + \nu |N_u|} \quad (\text{B.17})$$

Reescribiendo (B.3):

$$P(n(u)|n(v), v \in N_u) = \frac{\exp(-U_u(n(u)))}{\int_{-\infty}^{\infty} \exp(-U_u(x)) dx} \quad (\text{B.18})$$

$$= \frac{\exp(-U_u(n(u)))}{\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2} - B_x\right) dx} \quad (\text{B.19})$$

$$= \frac{\exp(-U_u(n(u)))}{\exp(-B_x) \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right) dx} \quad (\text{B.20})$$

$$= \frac{\exp(-U_u(n(u)))}{\exp(-B_x) \sqrt{2\pi}\sigma_x} \quad (\text{B.21})$$

Y la log-verosimilitud de ν :

$$l(\nu) = \log \left(\prod_u P(n(u)|n(v), v \in N_u) \right) \quad (\text{B.22})$$

$$= \sum_u \left(-U_u(n(u)) + B_x(u) - \log(\sigma_x(u)) - \log(\sqrt{2\pi}) \right) \quad (\text{B.23})$$

$$= \sum_u \left(-\frac{1}{2} n^2(u) - \nu \sum_{v \in N_u} (n(u) - n(v))^2 + \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2}{\frac{1}{2} + \nu |N_u|} \left(\sum_{v \in N_u} n(v) \right)^2 + \frac{1}{2} \log \left(\frac{1}{2} + \nu |N_u| \right) - \log(\sqrt{2\pi}) \right) \quad (\text{B.24})$$

Se busca calcular la esperanza de la expresión anterior:

$$= \mathbf{E} \left[-\frac{1}{2}n^2(u) - \nu \sum_{v \in N_u} (n(u) - n(v))^2 \right] \quad (\text{B.25})$$

$$+ \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2}{\frac{1}{2} + \nu|N_u|} \left(\sum_{v \in N_u} n(v) \right)^2 + \frac{1}{2} \log(1 + 4\nu) \right] \quad (\text{B.26})$$

$$= \mathbf{E} \left[-\frac{1}{2}n^2(u) - \nu \left(4n^2(u) - 2n(u) \sum_{v \in N_u} n(v) + \sum_{v \in N_u} n^2(v) \right) \right] \quad (\text{B.27})$$

$$+ \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2}{\frac{1}{2} + \nu|N_u|} \left(\sum_{v \in N_u} n(v) \right)^2 + \frac{1}{2} \log\left(\frac{1}{2} + 4\nu\right) \right] \quad (\text{B.28})$$

$$= \mathbf{E} \left[-\frac{1}{2}n^2(u) - 4\nu n^2(u) + 2\nu n(u) \sum_{v \in N_u} n(v) - \nu \sum_{v \in N_u} n^2(v) \right] \quad (\text{B.29})$$

$$+ \nu \sum_{v \in N_u} n^2(v) - \frac{\nu^2}{\frac{1}{2} + \nu|N_u|} \left(\sum_{v \in N_u} n(v) \right)^2 + \frac{1}{2} \log\left(\frac{1}{2} + 4\nu\right) \right] \quad (\text{B.30})$$

$$= \mathbf{E} \left[-\frac{1}{2}n^2(u) - 4\nu n^2(u) + 2\nu n(u) \sum_{v \in N_u} n(v) \right] \quad (\text{B.31})$$

$$- \frac{\nu^2}{\frac{1}{2} + \nu|N_u|} \left(\sum_{v \in N_u} n(v) \right)^2 + \frac{1}{2} \log\left(\frac{1}{2} + 4\nu\right) \right] \quad (\text{B.32})$$

luego,

$$S_1 = n^2(u) \quad (\text{B.33})$$

$$S_2 = n(u) \sum_{v \in N_u} n(v) \quad (\text{B.34})$$

$$S_3 = \left(\sum_{v \in N_u} n(v) \right)^2 \quad (\text{B.35})$$

Entonces, la expresión a maximizar es:

$$2\nu\bar{S}_2 - 4\nu\bar{S}_1 - \frac{\nu^2}{\frac{1}{2} + 4\nu}\bar{S}_3 + \frac{1}{2} \log\left(\frac{1}{2} + 4\nu\right) \quad (\text{B.36})$$

B.0.6. Solución

$$l(v) = 2\nu\bar{S}_2 - 4\nu\bar{S}_1 - \frac{\nu^2}{\frac{1}{2} + 4\nu}\bar{S}_3 + \frac{1}{2}\log\left(\frac{1}{2} + 4\nu\right) \quad (\text{B.37})$$

$$l(v) = a\nu - b\frac{\nu^2}{\frac{1}{2} + 4\nu} + \frac{1}{2}\log\left(\frac{1}{2} + 4\nu\right) \quad (\text{B.38})$$

$$l'(v) = a - b\frac{2\nu\left(\frac{1}{2} + 4\nu\right) - 4\nu^2}{\left(\frac{1}{2} + 4\nu\right)^2} + \frac{2}{\left(\frac{1}{2} + 4\nu\right)} \quad (\text{B.39})$$

$$l'(v) = a - b\frac{\nu + 4\nu^2}{\left(\frac{1}{2} + 4\nu\right)^2} + \frac{2}{\left(\frac{1}{2} + 4\nu\right)} \quad (\text{B.40})$$

$$l'(v) = \frac{a\left(\frac{1}{2} + 4\nu\right)^2 - b\nu - 4b\nu^2 + 2\left(\frac{1}{2} + 4\nu\right)}{\left(\frac{1}{2} + 4\nu\right)^2} \quad (\text{B.41})$$

$$l'(v) = \frac{\frac{a}{4} + 4a\nu + 16a\nu^2 - b\nu - 4b\nu^2 + 1 + 8\nu}{\left(\frac{1}{2} + 4\nu\right)^2} \quad (\text{B.42})$$

$$l'(v) = \frac{(16a - 4b)\nu^2 + (4a - b + 8)\nu + \frac{a}{4} + 1}{\left(\frac{1}{2} + 4\nu\right)^2} \quad (\text{B.43})$$

Sólo habría que resolver la ecuación cuadrática:

$$(16a - 4b)\nu^2 + (4a - b + 8)\nu + \frac{a}{4} + 1 = 0 \quad (\text{B.44})$$

Bibliografía

- Hervé Abdi. The Bonferonni and Sidák corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, 3:103–107, 2007.
- Milton Abramowitz y Irene A Stegun. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, tomo 55. Courier Corporation, 1972.
- Yoav Benjamini y Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, págs. 289–300, 1995.
- G. G. Berntson y J. T. Cacioppo. *Handbook of Neuroscience for the Behavioral Sciences*. John Wiley and Sons, 2009.
- Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, págs. 192–236, 1974.
- Julian Besag. Statistical analysis of non-lattice data. *The statistician*, págs. 179–195, 1975.
- Edward T Bullmore, John Suckling, Stephan Overmeyer, Sophia Rabe-Hesketh, Eric Taylor, y Michael J Brammer. Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *Medical Imaging, IEEE Transactions on*, 18(1):32–42, 1999.
- Arthur P Dempster, Nan M Laird, y Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, págs. 1–38, 1977.

- Karl J Friston, Keith J Worsley, RSJ Frackowiak, John C Mazziotta, y Alan C Evans. Assessing the significance of focal activations using their spatial extent. *Human brain mapping*, 1(3):210–220, 1994.
- Robert M Haralick, Stanley R Sternberg, y Xinhua Zhuang. Image analysis using mathematical morphology. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (4):532–550, 1987.
- Satoru Hayasaka y Thomas E Nichols. Validating cluster size inference: random field and permutation methods. *Neuroimage*, 20(4):2343–2356, 2003.
- Yosef Hochberg y Ajit C Tamhane. Multiple comparison procedures. 2009.
- Stefan J Kiebel, Jean-Baptiste Poline, Karl J Friston, Andrew P Holmes, y Keith J Worsley. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *Neuroimage*, 10(6):756–766, 1999.
- Jose L Marroquin, Rolando J Biscay, Salvador Ruiz-Correa, Alfonso Alba, Roxana Ramirez, y Jorge L Armony. Morphology-based hypothesis testing in discrete random fields: A non-parametric method to address the multiple-comparison problem in neuroimaging. *Neuroimage*, 59(3):3061–3074, 2012.
- T.E. Nichols y S. Hayasaka. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12:419–446, 2003.
- Patrick L Purdon y Robert M Weisskoff. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Human brain mapping*, 6(4):239–249, 1998.
- Roxana Ramírez. *Métodos para la detección de activaciones en campos aleatorios discretos*. Centro de Investigación en Matemáticas, 2011.
- Geraint Rees, T. Shepherd, B. Vemuri, S. Blackband, y T. Mareci. Single subject epoch (block) auditory fMRI activation data. 2015. URL <http://www.fil.ion.ucl.ac.uk/spm/data/auditory/>.

-
- Mariano Rivera, Omar Ocegueda, y Jose L. Marroquin. Entropy-Controlled Quadratic Markov Measure Field Models for Efficient Image Segmentation. *IEEE Transactions on Image Processing*, 16(12):3047–3057, 2007.
- Jean Serra. *Image analysis and mathematical morphology, v. 1*. Academic press, 1984.
- Stephen M Smith y Thomas E Nichols. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, 44(1):83–98, 2009.
- Keith J Worsley, Alan C Evans, S Marrett, y P Neelin. A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918, 1992.
- Eric Zarahn, Geoffrey K Aguirre, y Mark D'Esposito. Empirical analyses of BOLD fMRI statistics. *NeuroImage*, 5(3):179–197, 1997.
- Hui Zhang, Thomas E Nichols, y Timothy D Johnson. Cluster mass inference via random field theory. *Neuroimage*, 44(1):51–61, 2009.