Detection and localization of peaks in a smooth random field

Samuel Davenport (work with Thomas E. Nichols and Armin Schwartzman)

University of California, San Diego

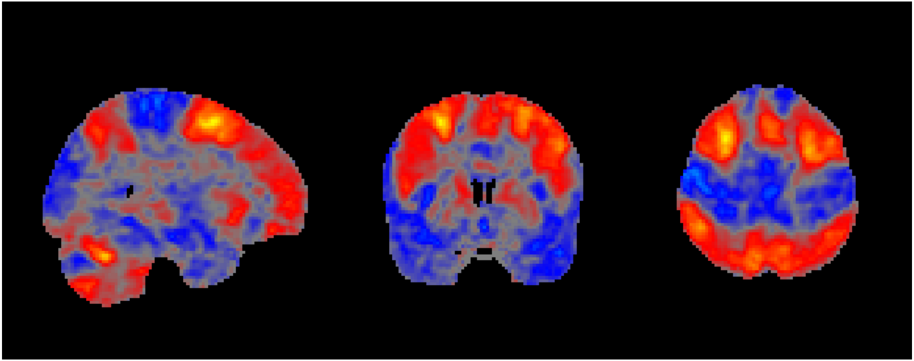
August 24, 2022

1. Conditions for Derivative Exchangeability
2. Main Theory
3. Results
4. Peaks of Cohen's d

References40

Peak inference

- In the era of large sample sizes the whole of the brain is found to be significant. Instead of detecting areas of activation we may want to perform more precise inference.
- In this presentation we will discuss how to provide confidence regions for peak location.



Notation

- Let $(Y_n)_{n\in\mathbb{N}}$ to be i.i.d almost surely differentiable random fields on an open domain $S \subset \mathbb{R}^D$.
- Let $\hat{\mu}_N = \frac{1}{N}\sum_{n=1}^{N} Y_n$ and $\hat{\sigma}_N^2 = \frac{1}{N-1}\sum_{n=1}^{N}(Y_n - \hat{\mu}_N)^2$.
- Let $T_N = \frac{\sqrt{N}\hat{\mu}_N}{\hat{\sigma}_N}$ be the t-statistic.
- Given a differentiable function $f : S \rightarrow \mathbb{R}^{D'}$, for $s \in S$, we shall write $\nabla f(s) \in \mathbb{R}^{D' \times D}$ to denote the gradient of f at s and use $\nabla^T f(s)$ to denote $(\nabla f(s))^T$.

Critical points

Let $f : S \to \mathbb{R}$ be twice differentiable.

Definition

We say that $s \in S$ is a **critical point** of f if $\nabla f(s) = 0$. Given a critical point s, we define s to be a **local maximum** of f if there is some $r > 0$ such that $f(s) = \sup_{t \in B_r(s)} f(t)$ and call a local maximum s **non-degenerate** if $\nabla^2 f(s) \prec 0$.

Local minima (and their non-degeneracy) can be defined similarly.

Conditions for Derivative Exchangeability

Derivative Exchangeability

In what follows we will want to be able to exchange expectation and differentiation. To do so:

Definition

We say that a random field $f : S \longrightarrow \mathbb{R}^{D'}$, some $D' \in \mathbb{N}$, is L_1-**Lipschitz at s** $\in S$ if there exists an integrable real random variable L and some ball $B(s) \subset S$ centred at s such that

$$\|f(t) - f(s)\| \leq L\|t - s\| \text{ for all } t \in B(s).$$

- This definition extends to subsets of S.
- This condition is useful because it implies that we can exchange the integral and the derivative.

DE condition

We say that a differentiable random field f on S satisfies the **DE (derivative exchangeability) condition** at $s \in S$ if $\mathbb{E}[f(t)]$ is differentiable at $t = s$ and

$$\mathbb{E}[\nabla f(t)] = \nabla \mathbb{E}[f(t)]$$

Lemma

Let $f : S \to \mathbb{R}^{D'}$ be an a.s. differentiable random field that is L_1-Lipschitz at $s \in S$. Then f satisfies the DE condition at s.

Sufficient conditions for Lipshitzness

Lemma

Let f be a random field on S which is a.s. differentiable on some ball $B(s) \subset S$, centred at $s \in S$. If $\mathbb{E} \sup_{t \in B(s)} \|\nabla f(t)\| < \infty$ then f is L_1-Lipschitz at s.

Proposition

Suppose that $f : S \to \mathbb{R}$ is an a.s. C^1 Gaussian random field. Then, for all $k \in \mathbb{N}$, $\mathbb{E} \sup_{t \in B(s)} \left\| \nabla f(t)^k \right\| < \infty$. Thus f^k is L_1-Lipschitz on S and therefore satisfies the DE condition on S.

Main Theory

Signal plus noise model

We assume a signal plus noise model:

$$\hat{\gamma}_N = \gamma + \eta_N$$

where $\eta_N \overset{\mathbb{P}}{\Longrightarrow} 0$ as $N \rightarrow \infty$.
This allows us to describe several scenarios of interest. E.g. the mean field:

$$\hat{\mu}_N = \mu + \frac{\sigma}{N}\sum_{n=1}^{N} \epsilon_n$$

and Cohen's d : by taking $\gamma = \frac{\mu}{\sigma}$ and $\eta_N = \left(d_N - \frac{\mu}{\sigma}\right)$. Where $d_N = \frac{\hat{\mu}_N}{\hat{\sigma}_N}$.

Assumptions on the signal

Assumption

- γ is C^2 and has $J \in \mathbb{N}$ critical points at locations $\theta_1, \dots, \theta_J \in S$, such that for $j = 1, \dots, J$ there exist non-overlapping compact balls $B_j \subset S$ such that $\theta_j \in \text{int}(B_j)$. Let $B_{\text{all}} = \bigcup_j B_j$ and assume that $C := \inf_{t \in S \setminus B_{\text{all}}} \|\nabla \gamma(t)\| > 0$.
- Let $P_{\max}$ be the subsets of $\{1, \dots, J\}$ corresponding to the non-degenerate local maxima of γ, respectively. Let $B_{\max} = \bigcup_{j \in P_{\max}} B_j$ and assume that

$$D_{\max} := -\sup_{t \in B_{\max}} \sup_{\|x\|=1} x^T \nabla^2 \gamma(t) x > 0.$$

Peak identifiability

Proposition

Suppose that $\nabla \eta_N \overset{\mathbb{P}}{\Longrightarrow} 0$, *and differentiable* $\gamma : S \rightarrow \mathbb{R}$ *which satisfies Assumption 1a. Suppose that for each* N, η_N *is a.s. differentiable, then as* $N \longrightarrow \infty$,

$$\mathbb{P}(\#\{t \in S \setminus B_{all} : \nabla \hat{\gamma}_N(t) = 0\} = 0) \longrightarrow 1.$$

Additionally assume that η_N *is a.s.* C^2 *with* $\nabla^2 \eta_N \overset{\mathbb{P}}{\Longrightarrow} 0$, *and let* $M_N = \left\{t \in S : \nabla \hat{\gamma}_N(t) = 0 \text{ and } \nabla^2 \hat{\gamma}_N(t) \prec 0\right\}$ *be the set of non-degenerate local maxima of* $\hat{\gamma}_N$. *Then, as* $N \longrightarrow \infty$, *for each* B_j *containing a non-degenerate local maximum of* γ:

$$\mathbb{P}(\#\{t \in M_N \cap B_j\} = 1) \longrightarrow 1.$$

CLT for the location of peaks of the mean

Theorem

For each $j = 1, \ldots, J$ corresponding to a local maximum of μ, let $\hat{\theta}_{j,n} = \text{argmax}_{t \in B_j} \hat{\mu}_N(t)$ (and for the minima let $\hat{\theta}_{j,N} = \text{argmin}_{t \in B_j} \hat{\mu}_N(t)$) and let $\hat{\boldsymbol{\theta}}_{\boldsymbol{N}} := (\hat{\theta}_{1,N}^T, \ldots, \hat{\theta}_{J,N}^T)^T$ and $\boldsymbol{\theta} := (\theta_1^T, \ldots, \theta_J^T)^T$. Then, under regularity assumptions on μ and the noise,

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_{\boldsymbol{N}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{A}^T)$$

as $N \longrightarrow \infty$. Here $\boldsymbol{A} \in \mathbb{R}^{DJ \times DJ}$ depends on $\nabla^2 \mu$ and $\boldsymbol{\Lambda} \in \mathbb{R}^{DJ \times DJ}$ depends on the covariance of ∇Y_1.

Proof idea Taylor expanding:

$$0 = \nabla \hat{\mu}_N(\hat{\theta}_{j,N}) = \nabla \hat{\mu}_N(\theta_j) + (\hat{\theta}_{j,N} - \theta_j)^T \nabla^2 \hat{\mu}_N(\theta_{j,N}^*) \tag{1}$$

Asymptotic Confidence Regions

For the jth peak let

$$\Sigma_j = (\nabla^2 \mu(\theta_j))^{-1} \text{cov}(\nabla^T Y_1(\theta_j))(\nabla^2 \mu(\theta_j))^{-1}$$

be the jth covariance. Then by the Theorem,

$$\sqrt{N}\Sigma_j^{-1/2}(\hat{\theta}_{j,N} - \theta_j) \sim \mathcal{N}(0, I_D) \implies N(\hat{\theta}_{j,N} - \theta_j)^T \Sigma_j^{-1}(\hat{\theta}_{j,N} - \theta_j) \sim \chi^2_D.$$

Thus, letting $\chi^2_{D,1-\alpha}$ be the $1 - \alpha$ quantile of the χ^2_D distribution it follows that

$$\left\{\theta : N(\hat{\theta}_{j,N} - \theta)^T \hat{\Sigma}_j^{-1}(\hat{\theta}_{j,N} - \theta) < \chi^2_{D,1-\alpha}\right\} \tag{2}$$

an asymptotic $(1 - \alpha)\%$ confidence region for θ_j, where

$$\hat{\Sigma}_j = (\nabla^2 \hat{\mu}(\hat{\theta}_j))^{-1} \hat{\Lambda}(\hat{\theta}_j)(\nabla^2 \hat{\mu}(\hat{\theta}_j))^{-1}.$$

Results

Simulations Description

- Given a mean function add noise to it (with different settings). In each setting we run $n_{\text{sim}} = 5000$ simulations.
- Noise generated via stationary Gaussian random fields formed by smoothing Gaussian white noise with a Gaussian kernel with FWHM in $\{3, \ldots, 9\}$.

Narrow peak

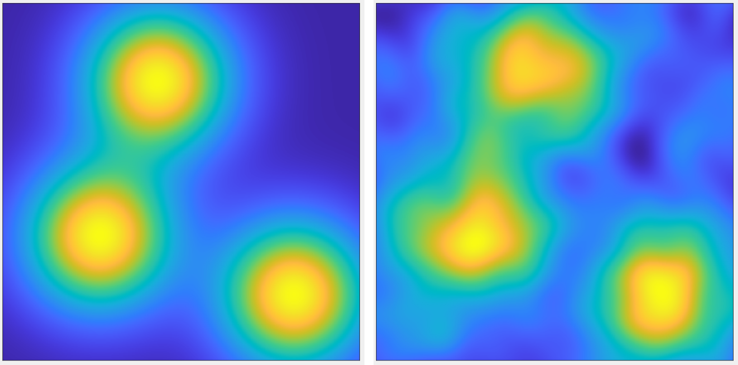


Figure 1: Left: True signal. Right: one realisation.

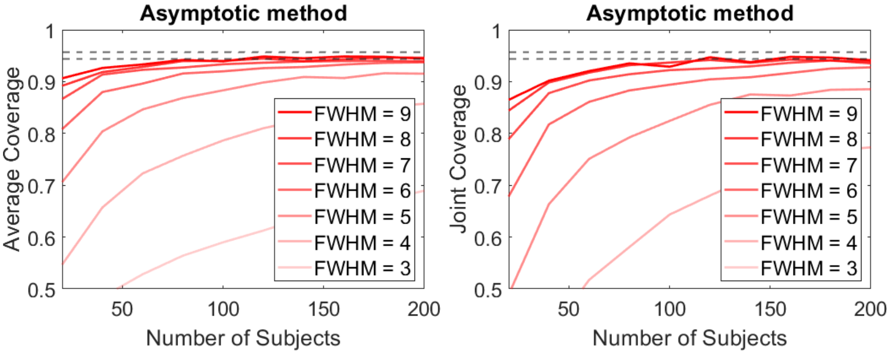
- For $\alpha \in (0,1)$, we define the **average empirical coverage** as

$$\frac{1}{J n_{\text{sim}}} \sum_{j=1}^{J} \sum_{i=1}^{n_{\text{sim}}} 1\left[\theta_j \in R_{i,j}^{\alpha}\right].$$

- We define the **empirical joint coverage** as

$$\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} 1\left[\theta_j \in R_{i,j}^{\alpha/J} \text{ for } 1 \leq j \leq J\right].$$

Comparing coverage rates



Wide peak

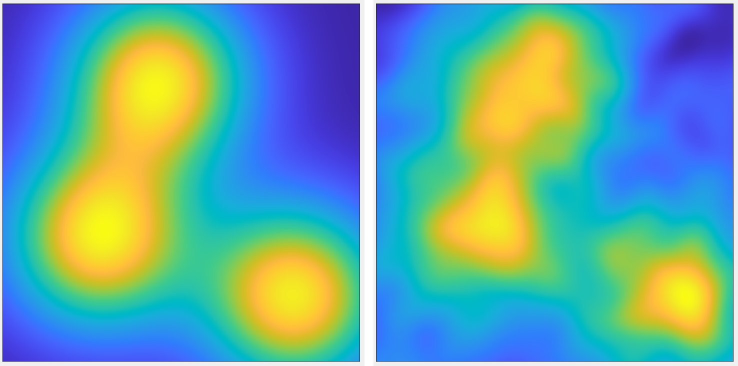
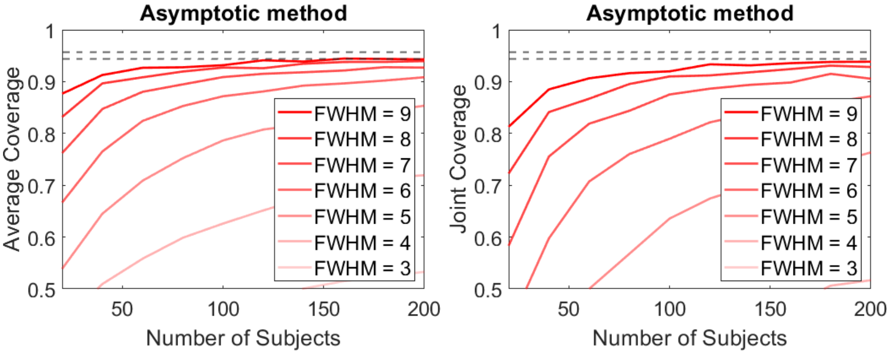


Figure 3: Left: True signal. Right: one realisation.

Comparing coverage rates



Accounting for the variance

$$0 = \nabla \hat{\mu}_N(\hat{\theta}_{j,N}) = \nabla \hat{\mu}_N(\theta_j) + (\hat{\theta}_{j,N} - \theta_j)^T \nabla^2 \hat{\mu}_N(\theta^*_{j,N}) \tag{3}$$

by Taylor, so rearranging,

$$\hat{\theta}_{j,n} - \theta_j = -\big(\nabla^2 \hat{\mu}_n(\theta^*_{j,n})\big)^{-1} \nabla^T \hat{\mu}_n(\theta_j)$$

Approximation of $(\nabla^2 \mu(\theta_j))^{-1}$ by $\left(\nabla^2 \hat{\mu}_n(\hat{\theta}_{j,n})\right)^{-1}$ leads to undercoverage as not all of the variance is accounted for since $\nabla^2 \hat{\mu}_n(\theta^*_{j,n})$ is a random variable. Instead,

$$\begin{aligned}\hat{\theta}_{j,n} - \theta_j &= -\left(\nabla^2 \hat{\mu}_n(\theta_j) + \frac{1}{2}(\hat{\theta}_{j,n} - \theta_j)^T \nabla^3 \hat{\mu}_n(\tilde{\theta}_{j,n})\right)^{-1} \nabla^T \hat{\mu}_n(\theta_j) \\ &\approx -\big(\nabla^2 \hat{\mu}_n(\theta_j)\big)^{-1} \nabla^T \hat{\mu}_n(\theta_j)\end{aligned}$$

Monte Carlo simulation from the distribution

We have

$$\begin{pmatrix} \nabla^T \hat{\mu}_n(\theta_j) \\ \mathbf{vech}(\nabla^2 \hat{\mu}_n(\theta_j)) \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \mathbf{vech}(\nabla^2 \mu_n(\theta_j)) \end{pmatrix}, \frac{1}{n} \begin{pmatrix} \Lambda & 0 \\ 0 & \Omega \end{pmatrix} \right)$$

and for $1 \leq k \leq K$ ($K \in \mathbb{N}$) we can approximate this by simulating from the following distribution

$$\begin{pmatrix} A_k \\ B_k \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \mathbf{vech}(\nabla^2 \hat{\mu}_n(\hat{\theta}_{j,n})) \end{pmatrix}, \frac{1}{n} \begin{pmatrix} \hat{\Lambda} & 0 \\ 0 & \hat{\Omega} \end{pmatrix} \right).$$

and calculating $\delta_{k,n} = (\mathbf{vech}^{-1}(B_{k,n}))^{-1} A_{k,n}$.

Monte Carlo confidence regions

Let $\hat{\Sigma}_j' = (\nabla^2 \hat{\mu}_n(\hat{\theta}_j))^{-1}\hat{\Lambda}(\nabla^2 \hat{\mu}_n(\hat{\theta}_j))^{-1}$ and for $0 < \alpha < 1$, choose λ_α such that

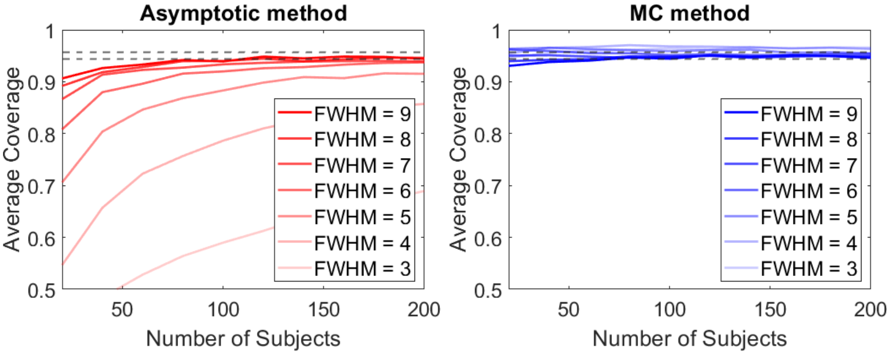
$$\frac{1}{K}\sum_{k=1}^{K} 1\left[n(\hat{\delta}_{k,n}^T(\hat{\Sigma}_j')^{-1}\delta_{k,n}) > \lambda_\alpha\right] = \frac{\lfloor \alpha K \rfloor}{K}.$$

Given this we define a $(1-\alpha)$ Monte Carlo confidence region to be

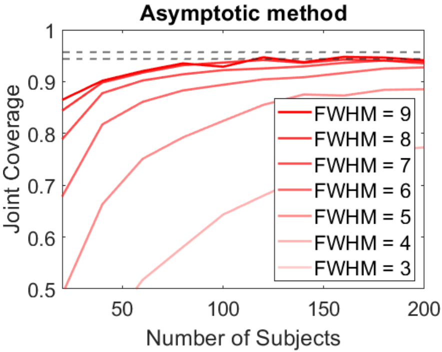
$$\left\{\theta : n(\hat{\theta}_{j,n} - \theta)^T(\hat{\Sigma}_j')^{-1}(\hat{\theta}_{j,n} - \theta) < \lambda_\alpha\right\}.$$

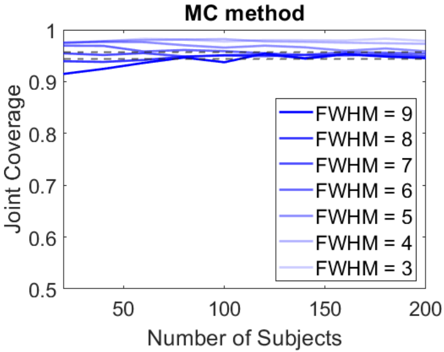
- These regions are asymptotically valid (for the same reason as the asymptotic cases)
- Under stationarity we can prove that these intervals are bigger than the asymptotic ones.

Comparing average coverage rates narrow signal

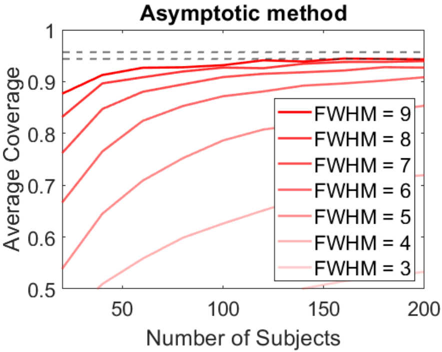


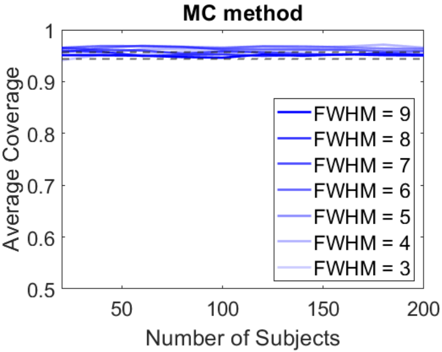
Comparing joint coverage rates narrow signal



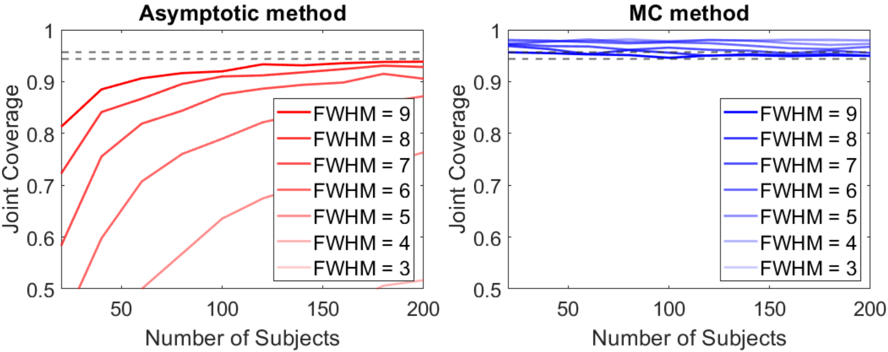


Comparing average coverage rates wide signal

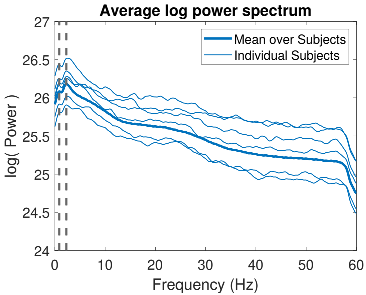




Comparing joint coverage rates wide signal



Application to MEG



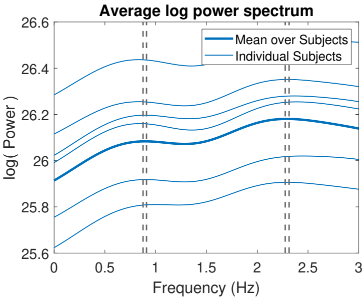


Figure 9: The top 2 peaks in the mean occur at 0.893 ± 0.017 Hz and 2.295 ± 0.019 Hz.

Note that in this case the asymptotic and MC confidence intervals are the same indicating that convergence has occurred.

Application to fMRI

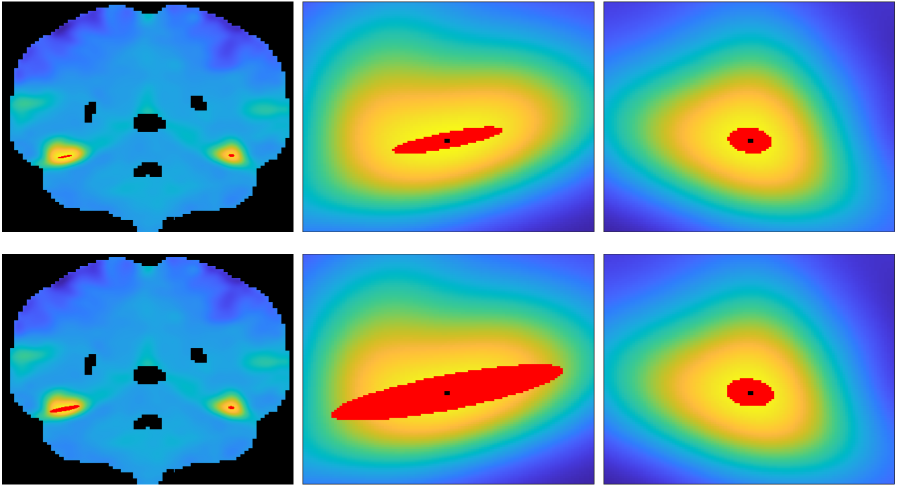


Figure 10: Peaks of the mean of 125 subjects

Peaks of Cohen's d

Asymptotic Results for Cohen's d

Recall that Cohen's d is

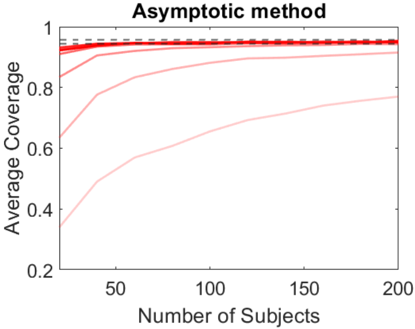
$$d_N = \frac{\hat{\mu}_N}{\hat{\sigma}_N}.$$

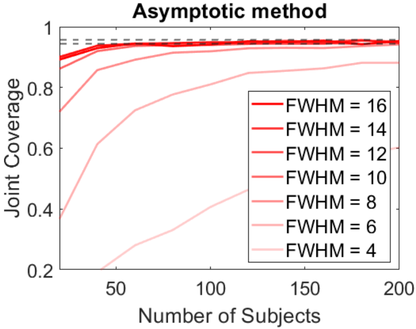
Then:

Theorem

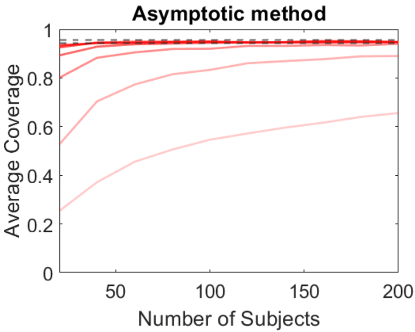
For each $j = 1, \ldots, J$ corresponding to a maximum of d, let $\hat{\theta}_{j,N} = \text{argmax}_{t \in B_j} d_N(t)$, (and for the minima let $\hat{\theta}_{j,N} = \text{argmin}_{t \in B_j} d_N(t)$) and let $\boldsymbol{\hat{\theta}_N} := (\hat{\theta}_{1,N}^T, \ldots, \hat{\theta}_{J,N}^T)^T$ and $\boldsymbol{\theta} := (\theta_1^T, \ldots, \theta_J^T)^T$. Then $\sqrt{N}(\boldsymbol{\hat{\theta}_N} - \boldsymbol{\theta})$ satisfies a CLT as $N \longrightarrow \infty$.

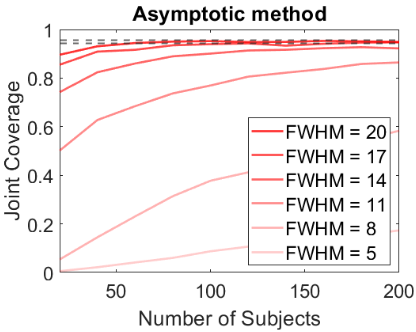
Cohen's d coverage - narrow peak





Cohen's d coverage - wide peak





Peaks of Cohen's d

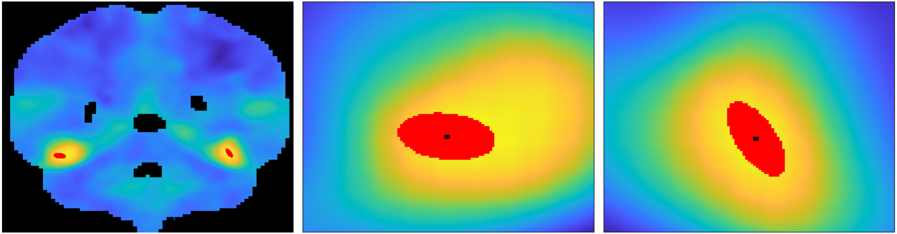


Figure 13: Peaks of Cohen's d of 125 subjects

Monte Carlo confidence intervals are difficult to derive for peaks of Cohen's d. Possible work for future research.

Conclusion

- The asymptotic confidence regions are valid in full generality and over multiple peaks.
- Under stationarity the Monte Carlo confidence regions provide substantially better counterparts.
- For this what is really needed is that the first and second derivatives are independent which is also guaranteed to also hold when the fields are constant variance.
- Local stationarity is probably sufficient.
- Asymptotic confidence for peaks of other statistics like R^2 etc should be possible to derive. Possibly Monte Carlo ones as well though that is more tricky.

Software and pre-print

- Software (in MATLAB) to perform inference on random fields is available at the RFTtoolbox (github.com/sjdavenport/RFTtoolbox). (E.g. for LKC estimation, Peak Inference, Peak Height distribution, confidence regions)
- Slides available at `sjdavenport.github.io/talks`.
- Pre-print on peak confidence regions is available on arxiv (Davenport, Nichols, & Schwarzman, 2022).

Monte Carlo distribution improves the coverage

Theorem

Suppose that A and B are independent real valued random variables with well defined densities p_A and p_B which are symmetric about $\mathbb{E}[A]$ and $\mathbb{E}[B]$ respectively. Assume that $p_A(x)$ is decreasing for $x > 0$ and increasing for $x < 0$, B is positive and that $\mathbb{E}[|B|] < \infty$. Then for all $x > 0$,

$$\mathbb{P}\left(\frac{A}{\mathbb{E}[B]} > x\right) \leq \mathbb{P}\left(\frac{A}{B} > x\right).$$

Bibliography

Davenport, S., Nichols, T. E., & Schwarzman, A. (2022). Confidence regions for the location of peaks of a smooth random field. *arXiv preprint arXiv:2208.00251*.