



Published in final edited form as:

Stat Methods Med Res. 2013 August ; 22(4): . doi:10.1177/0962280212448970.

A Bayesian Non-Parametric Potts Model with Application to Pre-Surgical FMRI Data

Timothy D. Johnson¹, Zhuqing Liu¹, Andreas J. Bartsch², and Thomas E. Nichols³

¹Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109

²Department of Neuroradiology, University of Heidelberg, Heidelberg, DE 69120

³Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Abstract

The Potts model has enjoyed much success as a prior model for image segmentation. Given the individual classes in the model, the data are typically modeled as Gaussian random variates or as random variates from some other parametric distribution. In this manuscript we present a non-parametric Potts model and apply it to an FMRI study for the pre-surgical assessment of peritumoral brain activation. In our model we assume that the Z-score image from a patient can be segmented into activated, deactivated and null classes, or states. Conditional on the class, or state, the Z-scores are assumed to come from some generic distribution which we model non-parametrically using a mixture of Dirichlet process priors within the Bayesian framework. The posterior distribution of the model parameters is estimated with a Markov chain Monte Carlo algorithm and Bayesian decision theory is used to make the final classifications. Our Potts prior model includes two parameters, the standard spatial regularization parameter and a parameter that can be interpreted as the a priori probability that each voxel belong to the null, or background state, conditional on the lack of spatial regularization. We assume that both of these parameters are unknown, and jointly estimate them along with other model parameters. We show through simulation studies that our model performs on par, in terms of posterior expected loss, with parametric Potts models when the parametric model is correctly specified, and outperforms parametric models when the parametric model is misspecified.

Keywords

Decision Theory; Dirichlet process; FMRI; hidden Markov random field; Non-parametric Bayes; Potts model

1 Introduction

Functional magnetic resonance imaging (FMRI) has found widespread use in the cognitive neurosciences due to its ability to non-invasively detect changes in brain activity induced by an experimental stimulus (1). The blood-oxygen-level dependent (BOLD) signal is the MRI contrast of blood deoxyhemoglobin and is a surrogate for neural activity. More recently, there has been a growing interest in using FMRI to assist in the pre-surgical planning of resection of brain tumors or epileptic foci (2). In these applications, the experimental paradigm used in the FMRI acquisition is carefully planned to detect changes in the BOLD signal near, around and within the tumor in an effort to assist the surgeon in excising the tumor while preserving brain tissue regions involved in common tasks, such as speaking, as determined by the FMRI data.

Our motivating data set comes from such a study. After a seizure with a speech arrest, a 32 year old female patient presented with a WHO grade II oligodendroglioma, a particular type

of a primary brain tumor derived from oligodendrocytes, in the left inferior frontal and insular lobe. Pre-surgery, the patient underwent an FMRI exam to determine vital peritumoral regions of the brain responsible for executive language functions and expressive speech. The patient was asked to recite tongue-twisters followed by periods of rest because she exhibited slight impairments in the repetition of phonemically challenging phrases while other speech and language functions were not disturbed. A high-resolution FLAIR (fluid attenuated inversion recovery) (3) sagittal image of the left hemisphere is displayed in the top row of Figure 1 where the tumor is clearly visible.

In a standard FMRI analysis a mass-univariate model is fit (e.g. with SPM (4) or FSL (5; 6)), where a regression model is fit independently at every volume element, or voxel. The resulting image of statistics is thresholded such that the risk of false positives over the brain is controlled at a specified level. The most common approach is to use a threshold that controls the familywise error rate using random field theory (RFT) (7). However, RFT rests on the assumption that the data arise as a realization of a smooth Gaussian random field. With no smoothing, FMRI data violate this assumption (8; 9) and thus the data are convolved with an isotropic Gaussian kernel prior to analysis to make this assumption hold, at least approximately. The result of this smoothing is that the activation (large, positive BOLD signal) and deactivation (large, negative BOLD signal) are spatially smeared out. Spatial diffusion of the signal is tolerable for the cognitive neurosciences as it increases the signal-to-noise ratio and gives results that may be typical for a population of subjects. In pre-surgical planning, however, spatial precision of the signal is paramount. According to electrical stimulation mapping (ESM), the size of speech-eloquent areas is known to vary between a few millimeters to a few centimeters. Thus, smoothing by a single uniform Gaussian kernel will inevitably smear out some activations and shrink or even extinguish others below the detection level. Both scenarios are highly undesirable for pre-surgical planning. Thus, we develop an image segmentation model that does not rely on such random field assumptions.

The Potts model (10) was originally developed as a generalization of the Ising model (11; 12) in the field of statistical mechanics. However, it has also been very successful as a prior model for image segmentation (12; 13; 14; 15). A crucial feature of the Potts model is that it does not smooth over abrupt changes in image intensity (12) as does the convolution of the image with an isotropic Gaussian kernel.

Several authors have proposed nonparametric hidden Markov random field models (such as the Potts model) (16; 17; 18; 19, e.g.). These authors all assume that the number of states in the hidden Markov random field is unknown and is to be estimated. To this end, they all use the Dirichlet process (20) which naturally clusters the data. These clusters become the states of the hidden Markov random field. Once the number of states is known, the image intensity at every voxel in a given state has the same (parametric) distribution. In our application, we know a priori that the BOLD activation data consists of exactly three states: activated, null and deactivated. Thus our model is different in that we assume that the number of states is known and model the image intensities of the voxels within each state non-parametrically. This is the key difference in our model from those previously proposed. Using one of the previously proposed models in our application would more than likely result in a model with more than three states and would be useless in identifying the activated, null and deactivated voxels.

This paper brings together several advanced statistical ideas including the Potts model, path sampling, mixture of Dirichlet process priors, slice sampling (via the Swendsen-Wang algorithm) and decision theory to estimate which voxels are deactivated, activated or null. Our simulation studies show that, in terms of minimizing the posterior expected loss, our

model performs equivalently to parametric Potts models when the parametric model is correctly specified and outperforms parametric models when the parametric model is misspecified.

The remainder of the paper is laid out as follows. In Section 2 we introduce notation and the Potts model along with our generalization to the Bayesian non-parametric Potts model (NP-Potts). Some of the key algorithmic details are outlined in Section 3. The motivating example is analyzed with our model in Section 4, where we contrast it to a standard Potts model, and present results from our simulation studies. We finish the paper with a discussion in Section 5.

2 Model

2.1 The Potts Model

Let \mathbf{Y} denote a digitized image of intensity values. Let \mathcal{S} denote the set of voxels (volume elements) if \mathbf{Y} is a three dimensional image or pixels (picture elements) if \mathbf{Y} is a two dimensional image. The cardinality of \mathcal{S} is N . The Potts model (10) relies on a neighborhood system and its associated cliques (13). Throughout we will assume a first-order neighborhood system. In the first-order neighborhood system, two pixels are neighbors if they share a common edge and two voxels are neighbors if they share a common face. Thus, each voxel in the interior of \mathbf{Y} has six neighbors. Henceforth, we refer only to voxels and reference them by a single index, say i . The correspondence between voxels and pixels should be obvious. We denote the fact that two voxels $i, i' \in \mathcal{S}$ are neighbors by $i \sim i'$. The set of neighbors of voxel $i \in \mathcal{S}$ is $\mathcal{N}_i = \{i' : i \sim i', i' \in \mathcal{S}\}$. For the first-order neighborhood system, the cliques are the singletons $\{i\}$ and all sets of pairs of neighbors $\{i, i'\}$. The image intensity, a random variable, at voxel i is Y_i with $\mathbf{Y} = (Y_1, \dots, Y_N)^T$. (We denote random variables by upper case letters and realized values by lower case letters).

Assume there exists a finite hidden Markov random field (21), $\mathbf{Z} = (Z_1, \dots, Z_N)^T$ on a finite state space \mathcal{C} with C states. It is easy to see that this random field forms a partition of the set of voxels \mathcal{S} with C equivalence classes (15). That is, voxels i and i' are equivalent if $Z_i = Z_{i'}$. For fMRI images we partition voxels into three equivalence classes denoted as deactivated, null and activated and will label these classes $-1, 0$, and 1 respectively: $\mathcal{C} = \{-1, 0, 1\}$. We will use the words ‘state’ and ‘class’ interchangeably. We denote the class to which voxel i belongs by $c(i)$. We consider the joint distribution of \mathbf{Y} and \mathbf{Z} given two sets of parameters: ϕ and θ :

$$\Pr(\mathbf{Y} \in A, \mathbf{Z} = \mathbf{z} | \phi, \theta) = \Pr(\mathbf{Y} \in A | \mathbf{Z} = \mathbf{z}, \phi) \Pr(\mathbf{Z} = \mathbf{z} | \theta)$$

where $A \subseteq \mathbb{R}^N$ is a measurable set. We assume \mathbf{Z} follows a Potts distribution, a special case of a Gibbs distribution, see reference (13), given by

$$\Pr(\mathbf{Z} = \mathbf{z} | \theta) = c^{-1}(\theta) \exp[-\mathcal{E}(\mathbf{z}; \theta)], \quad (1)$$

where $\mathcal{E}(\mathbf{z}; \theta)$ is the energy function, $\mathcal{E} : \mathcal{C}^N \rightarrow \mathbb{R} \cup \{\infty\}$. The energy function is typically written as a sum of Gibbs potential functions indexed by cliques Q : $V_Q(\mathbf{z}; \theta)$ where $V_Q \geq 0$ if Q is not a clique and for all $\mathbf{z}, \mathbf{z}' \in \mathcal{C}^N$ and for cliques $Q \subseteq \mathcal{S}$, we have $\mathbf{z}(Q) = \mathbf{z}'(Q)$ $V_Q(\mathbf{z}; \theta) = V_Q(\mathbf{z}'; \theta)$ where for a set $A \subseteq \mathcal{S}$, $\mathbf{z}(A) = \{z_i : i \in A\}$ (see (13) for details).

In what follows, we let $\mathcal{C} = \{0, 1\}$. Our Potts distribution is

$$\Pr(\mathbf{Z}=\mathbf{z}|\beta_0, \beta_1)=c^{-1}(\beta_0, \beta_1)\exp\left(-\beta_0\sum_{i\sim j}\{1-\delta_j[c(i)]\}-\beta_1\sum_{i\in\mathcal{S}}|z_i|\right) \quad (2)$$

where the first summation is taken over all neighboring pairs and $\delta_k(A)$ is the Dirac measure:

$$\delta_k(A)=\begin{cases} 0 & : k \notin A \\ 1 & : k \in A \end{cases},$$

that is, $\delta_j[c(i)] = 1$ if and only if $z_i = z_j$ and zero otherwise. $\beta_0 = 0$ is called the spatial regularization parameter and controls the strength of association between neighboring voxels. When $\beta_0 = 0$, class membership of a voxel is independent of its neighbors and all configurations, \mathbf{z} , are equally probable. For $\beta_0 > 0$, the Potts distribution favors configurations where neighboring voxels are members of the same equivalence class. The larger β_0 , the stronger the association between neighbors. When $\beta_0 = 0$ the second summation in equation (2) has a nice interpretation: $\Pr(\mathbf{Z} = \mathbf{z} | \beta_0 = 0, \beta_1) = \prod_{i \in \mathcal{S}} \Pr(Z_i = z_i | \beta_0 = 0, \beta_1) = \prod_{i \in \mathcal{S}} \exp(-\beta_1 |z_i|)$. Thus, for $z_i \in \{-1, 0, 1\}$, $\Pr(Z_i = z_i | \beta_0 = 0, \beta_1) = \exp(-\beta_1 |z_i|)$. In particular, $\Pr(Z_i = -1 | \beta_0 = 0, \beta_1) = \Pr(Z_i = 1 | \beta_0 = 0, \beta_1) = \exp(-\beta_1)/(1 + 2 \exp(-\beta_1))$ and $\Pr(Z_i = 0 | \beta_0 = 0, \beta_1) = (1 + 2 \exp(-\beta_1))^{-1}$. That is, the prior probability of a voxel belonging to the null state given $\beta_0 = 0$ and β_1 is $(1 + 2 \exp(-\beta_1))^{-1}$ and the prior probability of a voxel belonging to either the deactivated or activated state is $\exp(-\beta_1)/(1 + 2 \exp(-\beta_1))$. Note that $\beta_1 \in \mathbb{R}$.

The normalizing constant in equation (2) is given by

$$c(\beta_0, \beta_1) = \sum_{\mathbf{z} \in \mathcal{Z}} \exp\left(-\beta_0\sum_{i\sim j}\{1-\delta_j[c(i)]\}-\beta_1\sum_{i\in\mathcal{S}}|z_i|\right), \quad (3)$$

where the outer sum is over all possible configurations: $\mathcal{Z} = \{\mathbf{z} : z_i \in \mathcal{C}, i \in \mathcal{S}\}$. This normalizing constant is computationally intractable, even for moderate sized images and classes. For example a small $4 \times 4 \times 4$ image and 3 equivalence classes results in $\mathcal{Z} = 3^{64} \approx 3.4 \times 10^{30}$ possible configurations. This fact complicates Bayesian computation unless one assumes β_0 and β_1 are known, a priori. However, it is possible to estimate β_0 and β_1 along with all other parameters (see Section 3.2).

We assume that conditional on \mathbf{Z} , the Y_i are independent:

$$\Pr(\mathbf{Y} \in A | \mathbf{Z}=\mathbf{z}, \phi) = \prod_{i \in \mathcal{S}} \Pr(Y_i \in A_i | Z_i=j, \phi)$$

where $A_j \subseteq \mathbb{R}$ and $A_1 \times A_2 \times \dots \times A_N = A$. Furthermore, the conditional distribution of each Y_i is independent and identically distributed (iid)

$$\begin{aligned} [Y_i | Z_i=j, \phi] &\sim F_j(\phi_j) \quad \text{with} \\ [\phi_j | \psi_j] &\sim G(\psi_j), \end{aligned} \quad (4)$$

where $\phi_j \in \Phi$ and the $F_j(\phi_j)$ and $G(\psi_j)$ are parametric distributions.

2.2 The Bayesian Non-Parametric Potts Model

The G_j given in equation (4) is a parametric distribution. To relax this assumption we will use a mixture of Dirichlet process (MDP) priors (22; 23; 24). The model defined in equation (4) is replaced by

$$\begin{aligned} [Y_i|Z_i=j, \phi_i] &\sim F_j(\phi_i) \\ [\phi_i|Z_i=j, G_j] &\sim G_j \\ [G_j|G_{j0}, \alpha_j] &\sim DP(\alpha_j, G_{j0}), \end{aligned} \quad (5)$$

In what follows, F_j is Gaussian and $\phi_i = \{\mu_i, \sigma_i^2\}$ where μ_i is the mean and σ_i^2 is the variance of Y_i given $Z_i = j$. There are two important differences to point out between the models in equations (4) and (5). First, in equation (5), each Y_i is only conditionally independent and not independent and identically distributed as in equation (4). This is reflected in the fact that each Y_i has its own parameter set ϕ_i . Second, the distribution of each ϕ_i is specified as a random measure G_j that follows a Dirichlet process (DP) with base measure G_{j0} and concentration parameter α_j . The concentration parameter α_j determines how concentrated G_j is about G_{j0} with $E(G_j) = G_{j0}$. Furthermore, in our work, the base measure G_{j0} is Gaussian, parametrized by its mean, μ_{j0} , and variance, σ_{j0}^2 . We note that the conditional prior of \mathbf{Z} is given in equation (2). In order to finish specifying our model, we assign prior distributions to $\alpha_j, \beta_0, \mu_{-10}, \mu_{00}, \mu_{10}, \sigma_{j0}^{-2}$ and $\beta_{\sigma j}$:

$$\begin{aligned} \alpha_j &\sim \text{Gamma}(a, b), \quad j \in \{-1, 0, 1\}, \\ \beta_0 &\sim \text{Gamma}(c, d), \\ \mu_{-10} &\sim \text{Uniform}(\min(Y), 0), \\ \mu_{00} &\sim \text{Uniform}(\min(Y), \max(Y)), \\ \mu_{10} &\sim \text{Uniform}(0, \max(Y)), \\ \sigma_{j0}^{-2} &\sim \text{Gamma}(f, \beta_{\sigma j}), \quad j \in \{-1, 0, 1\}, \\ \beta_{\sigma j} &\sim \text{Gamma}(g, h). \end{aligned}$$

To specify a prior on β_1 , we first place a prior on $\Pr(Z_i = 0 | \beta_0 = 0, \beta_1) = p_{\beta_0} = (1 + 2 \exp(-\beta_1))^{-1}$. Let $p_{\beta_0} \sim \text{Beta}(p_a, p_b)$. This induces a prior on β_1 with density

$$\pi(\beta_1 | p_a, p_b) \propto (1/[1+2\exp(-\beta_1)])^{p_a} (2\exp(-\beta_1)/[1+2\exp(-\beta_1)])^{p_b}, \quad \beta_1 \in \mathbb{R}.$$

For the mean of the base measure for the deactivated and activated states we assume that the mean must be negative, respectively, positive. This aids in identifying the model, yet still allows, for example, an activated voxel to have a negative intensity (of course its neighbors will more than likely be large positive numbers). The hyperprior parameters $a, b, c, d, f, g, h, p_a, p_b$ are assumed known and will be identified below in the application section. Now each class in the Potts model is modeled non-parametrically as opposed to having a parametric structure. Note we place a hyperprior distribution on the inverse scale parameter, σ_{j0}^{-2} of the gamma prior for σ_{j0}^{-2} . The number of components of the MDP is sensitive to both the precision parameter α_j and to the variance σ_{j0}^2 of the base measure. By placing priors on the α_j and on the σ_{j0}^2 , we allow the data to inform on the amount of precision in the Dirichlet process, the variance of the base measure, G_{j0} and, hence, the number of components needed to fit the distribution of each class in the Potts model. We note here that our interest is not in the number of components that make up each class, rather it is in minimizing the posterior

expected loss defined in Section 3.4. Simulation studies confirm that the posterior expected loss is insensitive to the prior distributions on μ_j and σ_{j0}^{-2} . The model specified in equation (5) is non-parametric in the sense that the number of parameters is not fixed. The number of parameters goes to infinity as the number of voxels goes to infinity (which can happen as the size of the voxel is reduced). It is a mixture model in that if we marginalize over the μ_j the distribution of Y_i is a mixture of sampling distributions $F_j(\phi_j)$ with mixing measure G_j :

$$[Y_i|Z_i=j, G_j]=\int F_j(\phi_j)dG_j.$$

Thus, our model allows much more flexibility in modeling the three classes of voxels.

3 Algorithmic details

In this section we highlight some of the algorithmic details of the MCMC sampling scheme. In particular, in Section 3.1 we discuss the Swendsen-Wang algorithm (25) that we use to efficiently update the hidden Markov random field, \mathbf{Z} ; in Section 3.2 we discuss the path sampling algorithm (26) we use to estimate the ratio of normalizing constants we need in updating the parameters in the Potts model; in Section 3.3 we discuss the algorithm we use to update the mixture of Dirichlet process priors parameters (27); and in Section 3.4 we discuss the loss function we adopt in making our final assignment of voxels to classes using a Bayesian decision theoretic approach.

3.1 Swendsen-Wang algorithm

The Swendsen-Wang algorithm is a particular case of the slice sampler (28) that we now present. First, the full conditional posterior of \mathbf{Z} is

$$\Pr(\mathbf{Z}=\mathbf{z}|\cdot) \propto \pi(\mathbf{y}|\mathbf{Z}, \phi) \times \exp \left(-\beta_0 \sum_{i \sim j} \{1 - \delta_j[c(i)]\} - \beta_1 \sum_{i \in \mathcal{S}} |z_i| \right). \quad (6)$$

Now define ‘‘bond variables’’, W_{ij} for each neighbor pair $i \sim j$, where W_{ij} is a real, non-negative random variable, as follows. Let $\mathbf{W} = \{W_{ij}: i \sim j\}$ and define the conditional distribution of each W_{ij} given \mathbf{Z} to be uniform and independent with density

$$\pi(W_{ij}|\mathbf{Z}) = \exp(\beta_0[1 - \delta_j[c(i)]]) I\{0 \leq W_{ij} \leq \exp(-\beta_0[1 - \delta_j[c(i)]])\}. \quad (7)$$

Hence,

$$\pi(\mathbf{W}|\mathbf{Z}) = \prod_{i \sim j} \exp(\beta_0[1 - \delta_j[c(i)]]) I\{0 \leq W_{ij} \leq \exp(-\beta_0[1 - \delta_j[c(i)]])\}. \quad (8)$$

Furthermore,

$$\Pr(\mathbf{Z}=\mathbf{z}|\mathbf{W}, \cdot) \propto \pi(\mathbf{y}|\mathbf{Z}, \phi) \exp \left(-\beta_1 \sum_{i \in \mathcal{S}} |z_i| \right) \prod_{i \sim j} I\{0 \leq W_{ij} \leq \exp(-\beta_0[1 - \delta_j[c(i)]])\}. \quad (9)$$

To sample from the joint posterior of \mathbf{Z} and \mathbf{W} , we iteratively sample between equations (8) and (9). Integrating this joint distribution with respect to \mathbf{W} results in the full conditional of

\mathbf{Z} given in equation (6). Sampling $\mathbf{W} | \mathbf{Z}$ is straightforward: sample each $W_{ij} | \mathbf{Z}$ from equation (7). To sample a new $\mathbf{Z} | \mathbf{W}$ we note the following:

$$\Pr(W_{ij} > \exp(-\beta_0) | \mathbf{Z}) = \int_{\exp(-\beta_0)}^{\exp(-\beta_0(1-\delta_j[c(i)]))} dW_{ij} > 0 \iff \delta_j[c(i)] = 1 \quad (\text{i.e. } z_i = z_j).$$

That is to say, if $W_{ij} > \exp(-\beta_0)$ then z_i and z_j are constrained to be equal and this occurs with probability $1 - \exp(-\beta_0)$ (see equation (7)). In other words, neighbors in the same equivalence class are “bonded” with probability $1 - \exp(-\beta_0)$. Thus, \mathbf{W} partitions \mathcal{S} into clusters of voxels that share the same activation state. We can think of this new partition as a refinement of the partition of equivalence classes where each new equivalence class is a cluster of voxels that share the same activation state. For more details consult (15). For any particular cluster K ,

$$\Pr(\mathbf{Z}_K = \phi | \cdot) \propto \prod_{i \in K} \pi(y_i | Z_i = j, \phi_i) \exp(-\beta_1 |j|), \quad j = -1, 0, 1$$

where $\mathbf{Z}_K = \mathbf{j}$ is shorthand notation for $Z_i = j$ for all $i \in K$. Hence, each cluster can be updated independently according to its conditional distribution. For a nice discussion on the Swendsen-Wang algorithm see Higdon (14).

We alternate between Swendsen-Wang updates of the state space map \mathbf{Z} and single voxel Gibbs updates, adopting the strategy recommended by Higdon (14) who found that this aids in mixing of the posterior of \mathbf{Z} , to ensure movement in large patches.

3.2 Estimation of the potential energy parameters, β_0, β_1

The full conditionals for β_0 and β_1 are, respectively,

$$\pi(\beta_0 | \mathbf{Z}, \beta_1) \propto c^{-1}(\beta_0, \beta_1) \exp\left(-\beta_0 \sum_{i \sim j} \{1 - \delta_j[c(i)]\}\right) \beta_0^{c-1} \exp(-\beta_0 d)$$

and

$$\pi(\beta_1 | \mathbf{Z}, \beta_0) \propto c^{-1}(\beta_0, \beta_1) \exp\left(-\beta_1 \sum_{i \in \mathcal{S}} |z_i|\right) (1/[1+2\exp(-\beta_1)])^{p_a} (\exp(-\beta_1)/[1+2\exp(-\beta_1)])^{p_b}$$

where $\alpha(\beta_0, \beta_1)$ is given in equation (3). These full conditionals do not have a closed form from which we can easily sample and so we turn to the Metropolis-Hastings algorithm (29). The issue with this algorithm and the full conditionals is that we must compute the ratio of these normalizing constants when attempting to update either β_0 or β_1 . However, as mentioned previously, these constants are analytically and computationally intractable for any reasonably sized image.

Gelman and Meng (26) demonstrate how log ratios of normalizing constants can be efficiently estimated using path sampling. This is feasible as long as one can efficiently sample from the Potts prior (i.e. the prior distribution of $\mathbf{Z} | \beta_0, \beta_1$). The Swendsen-Wang

algorithm can be used for this purpose by replacing the full conditional of \mathbf{Z} with its prior in the previous section.

Here we outline the general procedure given in Gelman and Meng (26) for our particular application. We first consider the ratio of normalizing constants when updating β_0 for fixed β_1 .

Let $\lambda_{\beta_1}(a, b) = \ln[\alpha(\beta_1, b)/\alpha(\beta_1, a)]$. Then,

$$\begin{aligned} \lambda_{\beta_1}(a, b) &= \int_a^b \frac{\partial \ln[\alpha(x, \beta_1)]}{\partial x} dx = \int_a^b \frac{\partial \ln \sum_{\mathbf{z} \in \mathcal{Z}} \exp\left(-x \sum_{i \sim j} \{1 - \delta_j[c(i)]\} - \beta_1 \sum_{i \in \mathcal{S}} |z_i|\right)}{\partial x} dx \\ &= \int_a^b \sum_{\mathbf{z} \in \mathcal{Z}} \frac{\sum_{i \sim j} \{1 - \delta_j[c(i)]\} \exp\left(-x \sum_{i \sim j} \{1 - \delta_j[c(i)]\} - \beta_1 \sum_{i \in \mathcal{S}} |z_i|\right)}{\sum_{\mathbf{z} \in \mathcal{Z}} \exp\left(-x \sum_{i \sim j} \{1 - \delta_j[c(i)]\} - \beta_1 \sum_{i \in \mathcal{S}} |z_i|\right)} dx \\ &= - \int_a^b \left[\sum_{\mathbf{z} \in \mathcal{Z}} \left(\sum_{i \sim j} \{1 - \delta_j[c(i)]\} \right) \Pr(\mathbf{Z} = \mathbf{z} | x, \beta_1) \right] dx \\ &= - \int_a^b \mathbb{E}^{\Pr(\mathbf{Z} = \mathbf{z} | x, \beta_1)} \left[\sum_{i \sim j} \{1 - \delta_j[c(i)]\} \right] dx. \end{aligned} \tag{10}$$

Likewise, if $\lambda_{\beta_0}(a, b) = \ln[\alpha(\beta_0, b)/\alpha(\beta_0, a)]$, then

$$\lambda_{\beta_0}(a, b) = - \int_a^b \mathbb{E}^{\Pr(\mathbf{Z} = \mathbf{z} | \beta_0, x)} \left[\sum_{i \in \mathcal{S}} |z_i| \right] dx. \tag{11}$$

For both $\lambda_{\beta_1}(a, b)$ and $\lambda_{\beta_0}(a, b)$ we have two integrals to evaluate. We approximate the inner integrals, or expectations, using MCMC and are precomputed and saved prior to sampling from the posterior. We draw from $\Pr(\mathbf{Z} = \mathbf{z} | \beta_0, \beta_1)$ via the Swendsen-Wang algorithm (see Section 3.1) replacing the full conditional of \mathbf{Z} with its prior and approximating the expectations on a grid of values for β_0 and β_1 . We let β_0 vary on $[0, 2]$ in increments of 0.01. We determine the grid points for β_1 by varying p_{β_0} on $[0.01, 0.99]$ in increments of 0.01 and setting $\beta_1 = -\ln[(1 - p_{\beta_0})/2p_{\beta_0}]$. The outer integrals are evaluated numerically using the trapezoidal rule. For values of a and b not on the grid for which the expectations are evaluated, we linearly interpolate.

3.3 Posterior estimate of the MDP parameters

Over the past ten to fifteen years, there has been an explosion of Bayesian nonparametric papers stemming from the seminal works by Ferguson (20) and Antoniak and Ferguson (22) on Dirichlet processes. The first algorithm developed to sample from a Dirichlet process was the Polya urn scheme by Blackwell (30). Since then, several algorithms have been developed including the stick-breaking algorithm (31) and the Chinese restaurant process (32). Neal (27) compared six existing algorithms and presented two new algorithms that sample from the posterior of a Dirichlet process mixture model with non-conjugate priors. We adopt algorithm 8 from his publication to update $\phi_i = \{\mu_i, \sigma_i^2\}$ and refer the interested reader to his paper.

The precision parameters, β_j , are updated by the method given in Escobar and West (23). In particular for each j , define a latent random variable $\beta_j \in [0, 1]$ by

$$[\eta_j | \alpha_j, k_j] \sim \text{Beta}(\alpha_j + 1, N_j),$$

where k_j is the number of components in class j and N_j is the number of voxels in class j . Then,

$$[\alpha_j | \eta_j, k_j] \sim \pi_{\eta_j} \text{Gamma}(a + k_j, b - \ln \eta_j) + (1 - \pi_{\eta_j}) \text{Gamma}(a + k_j - 1, b - \ln \eta_j)$$

where $\pi_{\eta_j} = (a + k_j - 1) / (a + k_j - 1 + N_j (b - \ln \eta_j))$. See Escobar and West (23) for full details.

3.4 Loss function

The goal of using fMRI as a pre-surgical tool is to accurately determine which voxels belong to which class. That is, to determine the equivalence classes. We will make such assignments using a Bayesian decision theoretic approach. The surgeon wants to avoid removing or cutting through regions of the brain that are responsible for vital tasks such as speech, for example. Hence, a loss function that asymmetrically treats the various types of losses that can occur may be advantageous. In our example, there are three types of losses. Misclassification of a null voxel, misclassification of a deactivated voxel and misclassification of an activated voxel. It may be more important to correctly classify voxels and regions of the brain that are activated during a speech related task, for example, than correctly classifying those regions that are not responsible for speech as null.

Let $w_i \in \{-1, 0, 1\}$ denote the true state of voxel i . Let $d_i \in \{-1, 0, 1\}$ denote our decision about the state of voxel i . That is, $w_i = -1$ if voxel i is truly deactivated, $w_i = 0$ if voxel i has no true BOLD signal change from resting state and $w_i = 1$ if voxel i is truly activated. Let c_1 and c_2 be positive weights and define $\mathbf{w} = (w_1, \dots, w_N)$ and $\mathbf{d} = (d_1, \dots, d_N)$. Let δ_{ij} denote the Kronecker delta function which is equal to one if $i = j$ and zero otherwise. Our loss function is

$$L(\mathbf{d}, \mathbf{w}) = c_1 \sum_{i:w_i=-1} (1 - \delta_{-1d_i}) + \sum_{i:w_i=0} (1 - \delta_{0d_i}) + c_2 \sum_{i:w_i=1} (1 - \delta_{1d_i}).$$

When $c_1 = c_2 = 1$, the loss is simply the number of incorrectly classified voxels. When the weights are less than one, more weight is given to incorrectly classified null voxels. When $c_1 > 1$ more weight is given to incorrectly classified deactivated voxels and when $c_2 > 1$ more weight is given to incorrectly classified activated voxels, relative to null voxels.

Now, let \mathcal{D} denote the set of all possible decisions: $\mathcal{D} = \{\mathbf{d} : d_i \in \{-1, 0, 1\}, i \in \mathcal{S}\}$. From the Bayesian perspective, the optimal decision, \mathbf{d}^* , minimizes the posterior expected loss:

$$E^{\text{Pr}(\mathbf{Z}=\mathbf{z}|\mathbf{Y})} L(\mathbf{d}, \mathbf{z}) \sum_{\mathbf{z} \in \mathcal{Z}} L(\mathbf{d}, \mathbf{z}) \text{Pr}(\mathbf{Z}=\mathbf{z}|\mathbf{Y}).$$

That is

$$E^{\Pr(\mathbf{Z}=\mathbf{z}|\mathbf{Y})}L(\mathbf{d}^*, \mathbf{z}) = \inf_{\mathbf{d} \in \mathcal{D}} E^{\Pr(\mathbf{Z}=\mathbf{z}|\mathbf{Y})}L(\mathbf{d}, \mathbf{z}).$$

Define $q_{ji} = \Pr(Z_i = j | Y)$ for $j \in \{-1, 0, 1\}$. Suppose we draw T samples from the posterior with $\mathbf{z}^{(t)}$ denoting the value of \mathbf{Z} from draw t . Then the posterior estimate of q_{ji} is

$$\widehat{q}_{ji} = \sum_{t=1}^T \delta_{jz_i^{(t)}} / T \text{ and an estimator of posterior expected loss is}$$

$$\begin{aligned} \widehat{E}^{\Pr(\mathbf{Z}=\mathbf{z}|\mathbf{Y})}L(\mathbf{d}, \mathbf{z}) &= \frac{1}{T} \sum_{t=1}^T L(\mathbf{d}, \mathbf{z}^{(t)}) \\ &= c_1 \sum_{i \in \mathcal{I}} \widehat{q}_{-1i}(1 - \delta_{-1d_i}) + \sum_{i \in \mathcal{I}} \widehat{q}_{0i}(1 - \delta_{0d_i}) + c_2 \sum_{i \in \mathcal{I}} \widehat{q}_{1i}(1 - \delta_{1d_i}) \end{aligned} \quad (12)$$

which achieves its minimum by setting $d_i = -1$ if $\widehat{q}_{-1i} > \max(\widehat{q}_{0i}/c_1, c_2\widehat{q}_{1i}/c_1)$, $d_i = 0$ if $\widehat{q}_{0i} > \max(c_1\widehat{q}_{-1i}, c_2\widehat{q}_{1i})$, and $d_i = 1$ if $\widehat{q}_{1i} > \max(\widehat{q}_{0i}/c_2, c_1\widehat{q}_{-1i}/c_2)$. Typically, one sets $c_1 = c_2$, but this is not necessary.

4 Data Analysis and Simulation Studies

In this section we present results of a data analysis from a patient with a low-grade primary brain tumor in the left inferior frontal and insular lobe. We also present results from simulation studies, demonstrating the robustness of our model.

4.1 Data Analysis

As a clinical tool, the goal of pre-surgical fMRI is to assist the neurosurgeon in the planning and intra-operative navigation of resections or biopsies of intra-axial lesions, e.g. primary brain tumors, while avoiding damage to critical cerebral regions such as those responsible for motor, speech and language, visual or memory functions. Figure 1, top row, shows four sagittal slices of the FLAIR image of our example of a 32 year-old right-handed female patient suffering from a histologically proven oligodendroglioma of the left inferior frontal and insular lobe. The tumor is clearly visible in Figure 1. The paradigm to evoke executive speech- and language-related fMRI activations in the patient's brain was designed as a simple box-car of alternating on/off-episodes lasting 30 seconds each. During the on-episodes, the patient was asked to silently repeat auditorily presented tongue-twisters. The task was chosen because the only slight speech- and language-related deficit of the patient consisted of a slightly impaired tongue-twister repetition. During the off-episodes, the patient was asked to silently repeat "tock-tock-tock" without interruption, i.e. an unchallenging word train. Standard preprocessing and analysis was applied using FSL 4.1 (5; 33). The resulting z-statistic image is our data and we fitted the data with our NP-Potts model. Note that the z-statistic image is not smoothed (convolved with an isotropic Gaussian kernel) as is typically performed prior to analyses via random field theory. Smoothing is incorporated in the model via the Potts prior. However, as mentioned in the introduction the model also respects boundaries in the image; i.e. performs segmentation) (12, Ch. 2).

We now finish the definition of our model by specifying the particular prior and hyperprior parameter values. For the Dirichlet process concentration parameters, $\beta_j, j = -1, 0, 1$, we assign a gamma distribution with shape 3 and rate 2: $\beta_j \sim \text{Gamma}(3, 2)$. The hyperprior distributions for the base measures means are defined in Section 2.2 and the precisions, $\sigma_{j0}^{-2} \sim \text{Gamma}(3, \beta_{\sigma_j})$ with $\beta_{\sigma_j} \sim \text{Gamma}(1, 1)$, for $j \in \{-1, 0, 1\}$. For the spatial regularization parameter we assume $\beta_0 \sim \text{Gamma}(0.001, 0.001)$ —a rather uninformative prior. For β_1 we place an informative prior on p_{β_1} , which induces a prior on β_1 . fMRI

experiments are designed to find activation in very focal, specific regions of the brain. Thus, in virtually all fMRI experiments, a great majority of the voxels are not activated and the BOLD signal change evoked by the paradigm is expected to be zero in these areas. Thus, a priori, we believe that each voxel has a probability of 0.95 of belonging to the background, or null region. Thus, $p_{j0} \sim \text{Beta}(0.95 * (0.2N), 0.05 * (0.2N))$, where the beta distribution parameters have a ‘prior sample size’ interpretation. Specifically, in our specification we are assuming a prior sample size of 20% of the number of voxels, N , with 95% of the prior sample size belonging to the null class (background) and 5% belonging to either the deactivated or activated state.

We fit our model to the three-dimensional $64 \times 64 \times 40$ (pixel size: $3.0\text{mm} \times 3.0\text{mm} \times 3.45\text{mm}$) z-statistic image and use Bayesian decision theory to decide which voxels are activated, deactivated and null. Our collaborators want to ensure that truly activated or truly deactivated voxels are not misclassified and so we agreed on setting $c_1 = c_2 = 4$ in the posterior expected loss function found in equation (12), expressing the belief that false negatives are 4-times as costly as false positives. Thus, we classify any voxel, i , with posterior probability of activation greater than 0.2 ($\widehat{q}_{1i} > 0.2$) as activated and with posterior probability of deactivation greater than 0.2 ($\widehat{q}_{-1i} > 0.2$) as deactivated. The results are shown in the last two rows of Figure 1 where we display the FLAIR images with gray-scale overlay (color overlay in electronic version) depicting the probability of activated voxels (second row) and deactivated voxels (third row).

Clinicians are mainly interested in detecting functional activations. However, deactivations are also depicted because i) the baseline condition comprised a low-level rhythmic fluency task for comparison and ii) brain tumors can cause paradoxical blood oxygenation changes (34). Therefore, bidirectional statistical assessments and classifications are mandatory for clinical decision-making by pre-surgical fMRI (35). In the case presented here, strong frontal and peritumoral activations relevant for resection planning and neuronavigation were delineated (Figure 1, second row).

As a comparison, in Figure 2 we show results of a “standard” Potts model: compare equation (2) with

$$\Pr(\mathbf{Z}=\mathbf{z}|\beta_0)=c^{-1}(\beta_0)\exp\left(-\beta_0\sum_{i\sim j}\{1-\delta_j[c(i)]\}\right), \quad (13)$$

where β_0 is a random quantity to be estimated along with all other parameters. Our “standard” Potts model is parametrically defined by

$$\begin{aligned} [Y_i|Z_i=0, \mu, \sigma^2] &\sim N(\mu, \sigma^2) \\ [Y_i|Z_i=-1, r_{-1}, s_{-1}] &\sim \text{Gamma}^*(r_{-1}, s_{-1}) \\ [Y_i|Z_i=1, r_1, s_1] &\sim \text{Gamma}(r_1, s_1). \end{aligned}$$

We will refer to this model as the normal-gamma Potts model. In the normal-gamma Potts model, only voxels with negative intensities are allowed to belong to the deactivated state and only voxels with positive intensities are allowed to belong to the activated state. Gamma^* is the “reflected” gamma distribution defined for non-positive random variables as follows: if $Y \in (-\infty, 0]$, then $(y|r_{-1}, s_{-1}) = |y|^{r_{-1}-1} \exp(-s_{-1}|y|)$. One last restriction is put on the gamma and reflected gamma distributions: there must exist a mode greater than 0. This restriction was suggested by Woolrich, et al. (36) as it results in a sensible “belief about

the expected shape of the activation and deactivation distributions with respect to the non-activation.” This implies that $r_1 > 1$ and $r_{-1} > 1$. This model is a slight modification from that published in (36), which is currently implemented in FSL. For the normal-gamma Potts model, we adopt improper and uninformative priors: $(\mu) \sim 1$, $(\sigma^2) \sim 1/\sigma^2$, $[r] \sim \text{Gamma}(0.0001, 0.0001)$ and $[s] \sim \text{Gamma}(0.0001, 0.0001)$ for $s = -1, 1$. The same loss function as with the NP-Potts model was used to determine activated, deactivated and null voxels.

There exists dramatic differences between the results of our NP-Potts model and the normal-gamma Potts model. The simple experimental task of covert tongue-twister repetitions, when contrasted to a lower-level rhythm/fluency condition, is not expected to be supported by vast portions of the brain like those depicted in Figure 2. Intra-operative mapping during awake craniotomy confirmed that these extensive activations were over-inclusive and contained a large proportion of false-positive voxels. The differences are largely attributable to the fact that the standard Potts model implicitly assumes that the a priori probability of belonging to any one class is $1/C$ conditional on $\theta_0 = 0$ —or, in our case, $1/3$. In our NP-Potts model we place a hyperprior distribution on the probability of belonging to background and this has a significant impact on the number of voxels that are classified into each of the three states. Histograms, for comparison, of the three states for both the NP-Potts and the normal-gamma Potts model are shown in Figure 3.

In the next section, we present results from simulation studies that demonstrate the robustness of our model compared to several parametric Potts models.

4.2 Simulation Studies

We conducted a comprehensive simulation study and present the results in this section. We note here that all comparisons use the Gibbs distribution found in equation (2) and not that in equation (13). Our main goals in the simulation study are a) to demonstrate the robustness of our model to model misspecification and b) to demonstrate that we can accurately estimate the Potts model parameters θ_0 and θ_1 . To demonstrate the robustness of our model we compare our model to several parametric Potts models and show that the NP-Potts model performs as well or better in terms of smaller posterior expected loss. To demonstrate that we can accurately estimate θ_0 and θ_1 we consider the root mean squared error, relative bias and coverage of the 95% credible intervals.

Model robustness—We simulate the hidden Markov random field, \mathbf{Z} , using the Swendsen-Wang algorithm at all combinations of $\theta_0 \in \{0, 0.15, 0.25, 0.5, 0.75, 1\}$ and $\theta_1 \in \{\ln(4), \ln(2), \ln(1)\}$ (corresponding to a marginal a priori probability of a voxel belonging to the background of $1/3$, $1/2$ and $2/3$, respectively, assuming a three class state space). For \mathbf{Z} at each combination of θ_0 and θ_1 , we create an image \mathbf{Y} by populating each voxel i as determined by z_j from one of three distributions. For each voxel belonging to the null, or background, we draw independently from a $\mathcal{N}(0, 1)$ distribution. For voxels belonging to the deactivated or activated states, we draw independently from one of four distributions: a mixture of normals, a normal, a gamma and a log-normal. The specific distributions are given in Table 1 and were chosen such that segmentation is a difficult task (e.g. the resulting mixture distribution is unimodal). Each of these 4 models, which we call the normal-mixture (NM), the normal-normal (NN), the normal-gamma (NG) and the normal-log-normal (NL), generates data with a parametric structure. We then fit our model and three parametric Potts models to all simulated data sets and run each algorithm for 35000 iterations discarding the first 10000 iterations as burn-in. The three parametric Potts models we use to fit the data are the normal-gamma (background voxels fitted with a normal distribution, the activated voxels fitted with a gamma distribution and the deactivated voxels fitted with a reflected

gamma), the normal-normal model (all three states fitted with normal distributions) and the normal-log-normal model (background fitted with a normal distribution, activated voxels fitted with a log-normal distribution and deactivated voxels fitted with a reflected log-normal distribution). We then compare the posterior expected loss (with $c_1 = c_2 = 1$) between all models on all data sets. Figure 4 shows a bar plot of the posterior expected loss for all four models fitted to the four data sets generated from the four parametric Potts models. In this example $\beta_0 = 0.25$ and $\beta_1 = \ln(2)$. From this figure we see that the NP-Potts model performs as well (in terms of posterior expected loss) as the parametric models when the parametric model is correctly specified. When the parametric model is misspecified, the NP-Potts model outperforms (has smaller posterior expected loss) the parametric models. This trend is consistent for all simulation settings (results not shown) except when $\beta_0 = 0$. In Table 3 we present the proportion of correctly classified voxels from this simulation study for all three states as well as the total proportion of correctly classified voxels.

When $\beta_0 = 0$ there is no spatial correlation between voxels in the data. In this case, the analysis reduces to fitting a univariate distribution of image intensities with a mixture of distributions (one can think of modeling a univariate histogram with a mixture density). Obviously, in this case, a single mixture of Dirichlet process priors model suffices in fitting the distribution. The NP-Potts model breaks down as there is no information in the data or in the prior to inform how to fit the data with three non-parametric distributions and results are inconsistent when fitting the same data set multiple times. For our particular application (or for that matter, for any image analysis) this is not alarming to us as we expect spatial correlation between neighboring voxels. This raises the question of how much spatial regularization is necessary in the image data in order for our NP-Potts model to perform effectively. We found, through simulation studies, that a value of $\beta_0 = 0.15$, a rather small amount of spatial information, results in our NP-Potts model consistently performing as well or better than the parametric Potts models in terms of posterior expected loss.

Estimation of β_0 and β_1 —We consider the case when the true values are $\beta_0 = 0.5$ and $\beta_1 = \ln(2)$. We simulate 500 data sets using the NG model and 500 data set using the NM model (see above and Table 1). For all 1000 data sets we fit our NP-Potts model by drawing from the posterior 25000 times after a burn-in of 10000 draws. We consider the coverage of the 95% equal tail area credible intervals, the root mean squared error (RMSE) and the relative bias (rBias) of the posterior expected mean of β_0 and β_1 over the 1000 simulations. Results are given in Table 2. The coverage of the 95% credible intervals are near the nominal levels of 0.95. β_0 has a slight positive bias (relative bias 0.014) while the relative bias for β_1 is a bit larger in the negative direction (-0.074). However, in terms of p_{β} , which is arguably more interpretable, the relative bias is -0.026 , or about an average p_{β} of 0.487 over the simulations as opposed to the truth of 0.5. Overall, our model performs well and we can accurately estimate the Potts prior parameters. At the end of Section 2.2 we stated that the posterior expected loss is insensitive to the priors on the β_j and the σ_j^{-2} . For one randomly selected simulated data set from the normal-mixture model with true $\beta_0 = 0.5$ and $\beta_1 = \ln(2)$ we fit our model with different priors. For the β_j we consider the proposed prior, $\mathcal{G}(3, 2)$, and three other priors: $\mathcal{G}(1, 1)$, $\mathcal{G}(2, 4)$ and $\mathcal{G}(5, 1)$. The resulting posterior expected losses are 0.1515, 0.1521, 0.1518 and 0.1512, respectively. For σ_j^{-2} we consider the proposed prior, $\mathcal{G}(3, \beta_j)$ with $\beta_j \sim \mathcal{G}(1, 1)$ and two other priors: $\mathcal{G}(6, \beta_j)$ and $\mathcal{G}(1, \beta_j)$. The resulting posterior expected losses are, respectively, 0.1515, 0.1526 and 0.1527.

5 Discussion

We have presented a non-parametric Potts model and combined several ideas from the literature, including mixture of Dirichlet process priors, advanced sampling methods (path

sampling and slice sampling using the Swendsen-Wang algorithm) and decision theory to adapt loss to the pre-surgical planning context. Our simulation studies show that, in terms of the posterior expected loss, our model performs on par with parametric Potts models when the parametric model is correctly specified, and that it outperforms parametric models when the parametric model is incorrectly specified. Hence, we argue for the use of the NP-Potts model as one never knows the true distribution of the data. We have also shown that we can accurately estimate the Potts prior parameters β_0 and β_1 .

Computationally, 125,000 iterations of the MCMC algorithm takes about 3 hours of CPU time on a MacBook Pro with 8 GB of memory and a 3.06 GHz Intel Core 2 Duo processor. This is for the $64 \times 64 \times 40$ image in our data example. In this example, the number of voxels within the brain is 46999. Larger images will take more computation time.

In Section 3.2 we show how the Potts prior parameters, β_0 and β_1 , can be estimated using the idea of path sampling (26). The expectations in equations (10) and (11) have to be precomputed via MCMC simulation and the computation of these expectations is expensive. Furthermore, these expectations need to be computed for every patient since the size and shape of brains differ from patient to patient, making this implementation of the algorithm arduous and not very appealing. To circumvent this bottleneck, one could require that the Z-score maps of every patient be mapped onto a common atlas, compute the expectations once, off-line, on the atlas. The expectations could then be used for all patients. The downside of this is that the data are interpolated and this induces some smoothness in the interpolated image, which may not be desirable. It would be interesting to investigate the effects of this smoothing on parameter estimates and the posterior expected loss. Another possible alternative would be to approximate the likelihood with the pseudolikelihood (37). However use of the pseudolikelihood results in overestimation of the smoothing parameter and tends to oversmooth the data (38)—which is not a desirable outcome for presurgical planning, as argued earlier.

We feel it is imperative to always include the spatial regularization parameter β_0 in equation (2). For without it, one is assuming that all voxels are independent of one another and this assumption is almost never satisfied in medical imaging. On the other hand, β_1 plays a big role in determining the number of voxels that are classified as active or inactive, as is evident in our data analysis.

Acknowledgments

Dr. Johnson is partially funded by the U.S. National Institutes of Health [P01-CA87634].

References

1. Huettel, SA.; Song, AW.; McCarthy, G., editors. *Functional Magnetic Resonance Imaging*. 2. Sinauer Associates; 2008.
2. Yeo DTB, Meyer CR, MPJ, Minecan DN, Sagher O, Kluin KJ, et al. Formulation of current density weighted indices for correspondence between functional MRI and electrocortical stimulation maps. *Journal of Clinical Neurophysiology*. 2008; 119:2887–2897.
3. Bakshi R, Ariyaratana S, Benedict RH, Jacobs L. Fluid-attenuated inversion recovery magnetic resonance imaging detects cortical and juxtacortical multiple sclerosis lesions. *Archives of Neurology*. 2001; 58(5):742–748. [PubMed: 11346369]
4. Friston KJ, Jezzard P, Turner R. Analysis of functional MRI time-series. *Human Brain Mapping*. 1994; 1:153–171.
5. Smith SM, Jenkinson M, Woolrich CF, Behren TEJ, Johansen-Berg H, Bannister PR, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*. 2004; 23(S1):208–219.

6. Smith SM, Beckmann CF, Ramnani N, Woolrich MW, Bannister PR, Jenkinson M, et al. Variability in fMRI: A re-examination of intersession differences. *Human Brain Mapping*. 2005; 24:248–257. [PubMed: 15654698]
7. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant voxels in images of cerebral activation. *Human Brain Mapping*. 1996; 4:48–73.
8. Petersson KM, Nichols TE, Poline JB, Holmes AP. Statistical limitations in functional neuroimaging II. Signal detection and statistical inference. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 1999; 354(354):1261–1281.
9. Nichols TE, Hayasaka S. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical Methods in Medical Research*. 2003; 12(5):419–446. [PubMed: 14599004]
10. Potts RB. Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophic Society*. 1952; 48:106–109.
11. Ising E. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*. 1925; 31(1):253–258.
12. Winkler, G. *Image Analysis, Random Fields and Markov Chain Monte Carlo Methods*. Springer; Berlin: 2003.
13. Brémaud, P. *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer; New York: 1999.
14. Higdon DM. Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association*. 1998; 93(442):585–595.
15. Johnson TD, Pierr M. A Bayesian analysis of dual autoradiographic images. *Computational Statistics and Data Analysis*. 2009; 53:4570–4583.
16. da Silva ARS. A Dirichlet process mixture model for brain MRI tissue classification. *Medical Image Analysis*. 2007; 2:169–182.
17. Orbanz P, Buhmann J. Nonparametric Bayes image segmentation. *International Journal of Computer Vision*. 2008; 77:25–45.
18. Du, L.; Ren, L.; Dunson, D.; Carin, L. A Bayesian model for simultaneous image clustering, annotation and object segmentation. In: Bengio, Y.; Schuurmans, D.; Lafferty, J.; Williams, CKI.; Culotta, A., editors. *Advances in Neural Information Processing Systems*. Vol. 22. MIT Press; 2009. p. 486-494.
19. Chatzis SP, Tsechpenakis G. The infinite hidden Markov random field model. *IEEE Transactions on Medical Imaging*. 2010; 21(6):1004–1014.
20. Ferguson TS. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*. 1973; 1(2):209–230.
21. Besag J. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society Series B*. 1974; 6:192–236.
22. Antoniak CE. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The Annals of Statistics*. 1974; 2(6):1152–1174.
23. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*. 1995; 90(430):577–588.
24. Müller P, Quintana FA. Nonparametric Bayesian data analysis. *Statistical Science*. 2004; 19(1):95–110.
25. Swendsen RH, Wang JS. Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*. 1987; 58:86–88. [PubMed: 10034599]
26. Gelman A, Meng X. Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statistical Science*. 1998; 13(2):163–185.
27. Neal R. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*. 2000 Jun; 9(2):249–265.
28. Robert, CP.; Casella, G. *Monte Carlo Statistical Methods*. Springer; New York: 2005.
29. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970; 57(1):97–109.

30. Blackwell D, MacQueen JB. Ferguson distributions via Polya urn schemes. *The Annals of Statistics*. 1973; 1(2):353–355.
31. Ishwaran H, James LF. Gibbs Sampling for Stick-Breaking Priors. *JASA*. 2001; 96:161–173.
32. Pitman J. Exchangeable and Partially Exchangeable Random Partitions. *Probability Theory and Related Fields*. 1995; 102(2):145–158.
33. Woolrich MW, Jbabdi SBP, Chappell MSM, Behrens T, et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*. 2009; 45:173–186.
34. Fujiwara N, Sakatani K, Katayama Y, Murata Y, Hoshino T, Fukaya C, et al. Evoked-cerebral blood oxygenation changes in false-negative activations in BOLD contrast functional MRI of patients with brain tumors. *NeuroImage*. 2004; 21(4):1464–1471. [PubMed: 15050571]
35. Bartsch AJ, Homola G, Biller A, Solymosi L, Bendszus M. Diagnostic functional MRI: illustrated clinical applications and decision-making. *Journal of Magnetic Resonance Imaging*. 2006; 23(6): 921–932. [PubMed: 16649199]
36. Woolrich MW, Behrens TEJ, Beckmann CF, Smith SM. Mixture Models With Adaptive Spatial Regularization for Segmentation With an Application to FMRI Data. *IEEE Transactions on Medical Imaging*. 2005; 24(1):1–11. [PubMed: 15638182]
37. Besag J. Statistical Analysis of Non-Lattice Data. *The Statistician*. 1975; 24(3):179–195.
38. Melas DE, Wilson SP. Double Markov random fields and Bayesian image segmentation. *IEEE Transactions on Signal Processing*. 2002; 50:357–365.

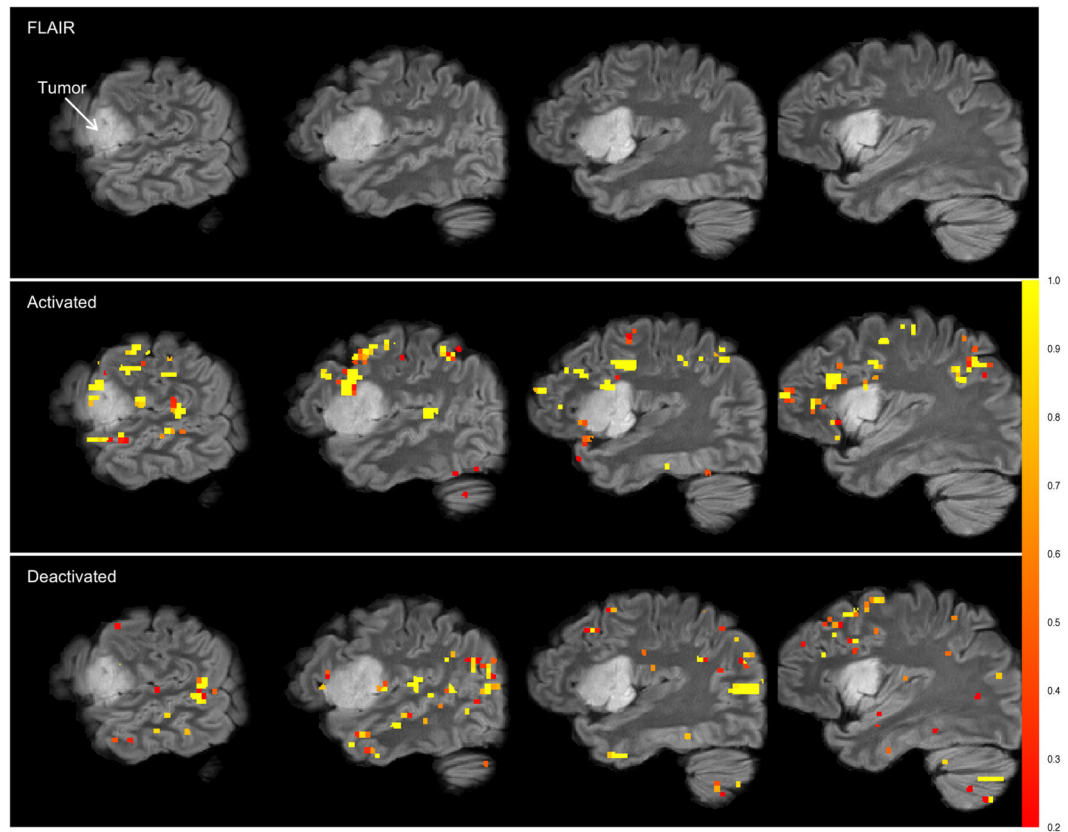


Figure 1.

Results from the NP-Potts model. Top row: Four sagittal slices of the high resolution FLAIR image in the left hemisphere. Second row: FLAIR image with activation overlay. Bottom row: FLAIR image with deactivation overlay. The gray-scale bar (color bar in electronic version) represents the posterior probability of both activation (for the second row) and deactivation (for the bottom row). The overlays are highly pixelated as the analysis is performed in the functional MRI space which has a much lower resolution than the high resolution FLAIR image. The hyperprior distribution on p_0 is $p_0 \sim (\text{Beta}(0.95(0.2N), 0.05(0.2N)))$. Loss function is (12) with $c_1 = c_2 = 4$.

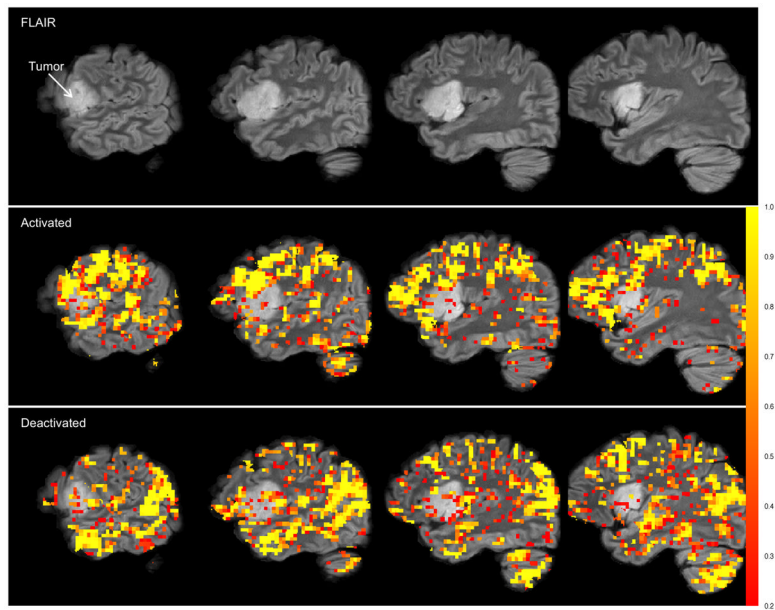


Figure 2.

Results from the normal-gamma model. Top row: Four sagittal slices of the high resolution FLAIR image in the left hemisphere. Second row: FLAIR image with activation overlay. Bottom row: FLAIR image with deactivation overlay. The gray-scale bar (color bar in electronic version) represents the posterior probability of both activation (for the second row) and deactivation (for the bottom row). The overlays are highly pixelated as the analysis is performed in the functional MRI space which has a much lower resolution than the high resolution FLAIR image. The prior on p_{j0} is fixed: $p_{j0} = 1/3$. Loss function is (12) with $c_1 = c_2 = 4$.

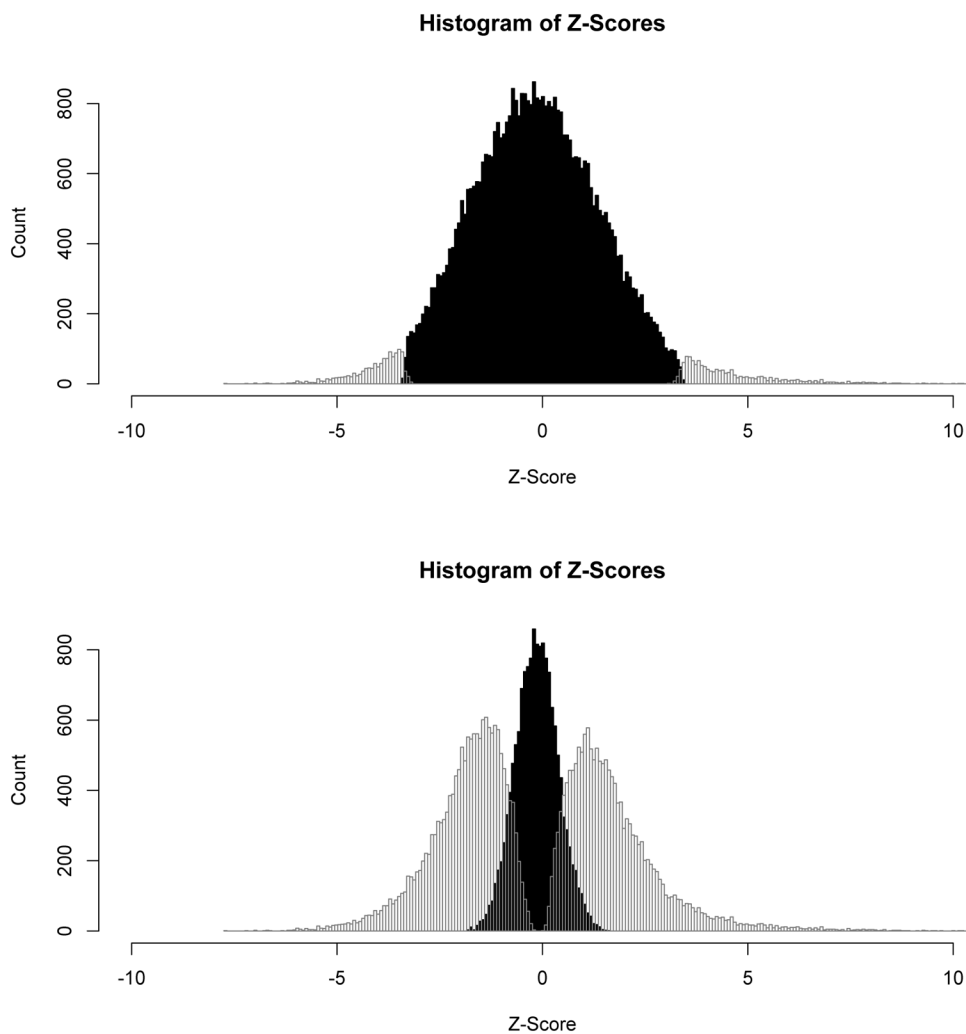


Figure 3. Histograms of the Z-scores from the final classifications into deactivated, null and activated states. The top histogram is from the DP-Potts model and the bottom histogram is from the normal-gamma model. The black histogram corresponds to the null classified voxels and their Z-scores, The outlined histograms to the left of the null histogram correspond to the deactivated voxels and those to the right, the activated voxels.

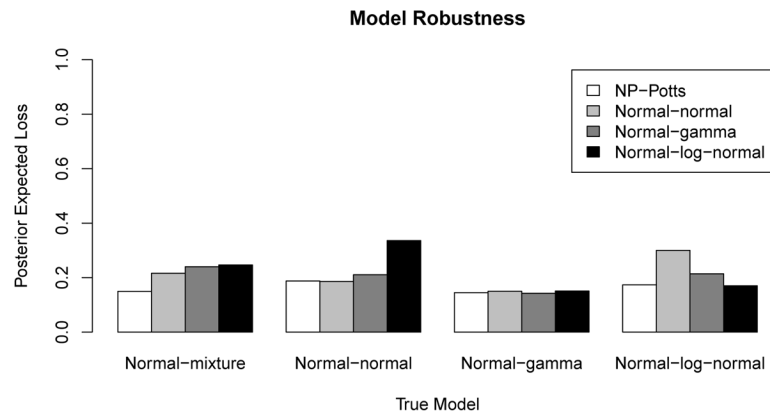


Figure 4. Robustness of the NP-Potts model. The NP-Potts model performs on par with the correctly specified model and performs better than misspecified models. $\theta_0 = 0.25$, $\theta_1 = \ln(2)$ (marginal prior probability = 0.5). The gray scale bars represent the posterior expected loss from the model used to fit the data (see legend) generated from the true model.

Table 1

The four parametric models used in the simulation studies. Gamma* and $\ln \mathcal{N}^*$ represent the reflected gamma and log-normal distributions.

	State		
	Deactivated	Null	Activated
Normal-mixture (NM)	$0.75\mathcal{N}(-2, 1) + 0.25\mathcal{N}(-5, 1)$	$\mathcal{N}(0, 1)$	$0.75\mathcal{N}(2, 1) + 0.25\mathcal{N}(5, 1)$
Normal-normal (NN)	$\mathcal{N}(-2, 1)$	$\mathcal{N}(0, 1)$	$\mathcal{N}(2, 1)$
Normal-gamma (NG)	Gamma*(5, 2)	$\mathcal{N}(0, 1)$	Gamma(5, 2)
Normal-log-normal (NL)	$\ln \mathcal{N}^*(0.75, 0.25)$	$\mathcal{N}(0, 1)$	$\ln \mathcal{N}(0.75, 0.25)$

Table 2Estimation of θ_0 and θ_1 in the simulation study.

	Parameter	
	θ_0 (truth: 0.5)	θ_1 (truth: ln(2))
RMSE	0.0167	0.067
rBias	0.014	-0.074
Coverage	0.964	0.957

Table 3

Proportion of correctly classified voxels from the simulation study. The results show that the NP-Potts perform equally to the correctly specified parametric model and outperforms the misspecified parametric models.

True Model	Fitted Model	Deactivated	Null	Activated	Overall
Normal-mixture	NP-Potts	0.680	0.931	0.680	0.845
	Normal-normal	0.346	0.997	0.343	0.774
	Normal-gamma	0.284	0.999	0.281	0.754
	Normal-log-normal	0.261	1.000	0.261	0.747
Normal-normal	NP-Potts	0.652	0.896	0.657	0.813
	Normal-normal	0.662	0.892	0.663	0.813
	NG	0.456	0.968	0.456	0.793
	NL	0.000	1.000	0.000	0.657
Normal-gamma	NP-Potts	0.734	0.914	0.757	0.857
	Normal-normal	0.667	0.936	0.710	0.851
	Normal-gamma	0.739	0.914	0.755	0.857
	Normal-log-normal	0.703	0.932	0.717	0.856
Normal-log-normal	NP-Potts	0.728	0.885	0.732	0.832
	Normal-normal	0.840	0.001	0.813	0.284
	Normal-gamma	0.381	0.986	0.371	0.777
	Normal-log-normal	0.746	0.879	0.734	0.831