# NOVEL PARAMETER PRIORS FOR BAYESIAN SIGNAL IDENTIFICATION

*Anthony Quinn*

Department of Electronic and Electrical Engineering, University of Dublin, Trinity College, Dublin 2. Ireland.
**Phone:** +353-1-6081863/6081580; **Fax:** +353-1-6772442; **Email:** aquinn@ee.tcd.ie

## ABSTRACT

The problem of eliciting priors on the parameter space of a signal hypothesis is considered in this paper, and two lesser-known approaches are emphasized. Each yields conservative priors appropriate for data-dominated Bayesian parameter inference. They are based, respectively, on the principles of (i) *a posteriori* transformation invariance, and (ii) *a priori* maximum entropy. Novel priors on a wide class of signal models are deduced. Their ability to regularize inference of the difference frequency between closely spaced tones is considered, and they are compared with the Ockham Prior which was studied in previous work.

## 1. INTRODUCTION

In the Bayesian approach, inferences concerning unknown parameters are drawn from the *a posteriori* density, which is formed, from Bayes' Theorem [1, 2], as the product of the *Likelihood Function (LF)* and the *prior*. It is the problem of assigning prior distributions in signal identification which is addressed in this paper. The prior not only quantifies any relevant information available before observations are made, but it can also be used to modify inferences so that they behave appropriately (Section 2.1).

The problem of prior elicitation remains the single most contentious area in all of mathematical inference. Many criteria have been proposed in the literature [1–7], some of which have been used in Bayesian signal identification:

(i) **Bayes' Postulate:** this states that a uniform prior should be assigned when 'no' prior information is available. Inferences are then drawn exclusively from the LF.

(ii) **Jeffreys' Priors:** $\Theta = \theta$ is a *location parameter* if the LF has the form

$$\mathrm{p}(x \mid \theta) = h(x - \theta).$$

In this case, Jeffreys' Prior for $\Theta$ is uniform, as in (i). $\Sigma = \sigma$ is a *scale parameter* if

$$\mathrm{p}(x \mid \sigma) = \frac{1}{\sigma} h(\frac{x}{\sigma}),$$

for which Jeffreys' Prior is $\mathrm{p}(\sigma) \propto 1/\sigma$. The former is employed widely as the prior for 'signal' parameters, the latter as the prior on the standard deviation of the additive 'noise' [1, 5, 8].

(iii) **Conjugate Priors:** the prior, $q(\theta)$, is conjugate to the LF, $\mathrm{p}(x \mid \theta)$, if the resulting posterior, $\pi(\theta \mid x)$, has the same form as $q(\cdot)$. This property is useful in ensuring tractability in problems of sequential inference [2].

The sequel defines the problem of assigning priors consistently (*i.e.* "elicitation"), and presents two strategies which are not widely known in the signal identification literature.

## 2. PRIOR ELICITATION

Consider a parametric hypothesis, $\mathcal{I}$, which seeks to explain a set of observations, $\mathbf{D} = \mathbf{d} \in \overline{\mathbb{C}}^N$, in terms of a set of unknown parameters, $\mathbf{\Theta} = \boldsymbol{\theta} \in \overline{\Theta} \subset \overline{\mathbb{C}}^k$, where $\overline{\mathbb{C}}$ denotes the set of complex numbers, $\overline{\Theta}$ is the space of $\boldsymbol{\theta}$, and notation of the type '$\mathbf{X} = \mathbf{x}$' signifies a realization, $\mathbf{x}$, for the probabilistic parameter (p.p.) [9] vector, $\mathbf{X}$. $\mathbf{d}$ may be viewed as a vector of $N$ discrete samples from a random process. The direct probability of $\mathbf{d}$ given candidate values of $\boldsymbol{\theta}$, *i.e.* $\mathrm{p}(\mathbf{d} \mid \boldsymbol{\theta}, \mathcal{I})$, may be interpreted as a function, $l(\boldsymbol{\theta} \mid \mathbf{d}, \mathcal{I})$, of $\boldsymbol{\theta}$ itself, being the *Likelihood Function (LF)*. In this paper, $\mathcal{I}$ additively decomposes $\mathbf{d}$ into a 'signal', $\mathbf{s} = \mathbf{g}(\boldsymbol{\theta})$, and 'noise', $\mathbf{e}$:

$$\mathbf{d} = \mathbf{g}(\boldsymbol{\theta}) + \mathbf{e}, \tag{1}$$

where

$$\mathbf{e} \sim \mathrm{p}_E(\mathbf{e} \mid \mathcal{I}). \tag{2}$$

'$\sim$' denotes 'is distributed as', and $\mathbf{g}(\cdot)$ is the sampled signal function, parameterized by $\boldsymbol{\theta}$. It will be assumed that the analytical form of this function is known, so that the signal identification problem reduces to one of estimating $\mathbf{\Theta}$. The framework is readily extended to the case of model uncertainty [5, 9]. From (1,2), the LF is given by

$$\mathrm{p}(\mathbf{d} \mid \boldsymbol{\theta}, \mathcal{I}) \equiv l(\boldsymbol{\theta} \mid \mathbf{d}, \mathcal{I}) = \mathrm{p}_E(\mathbf{d} - \mathbf{g}(\boldsymbol{\theta}) \mid \mathcal{I}). \tag{3}$$

Much of the *orthodox* signal identification literature is based on inferences from the LF. In contrast, the Bayesian approach recognizes that degrees of belief in a proposition are quantified by Probability. For example, the proposition that the (unknown) parameters, $\mathbf{\Theta}$, take on value $\boldsymbol{\theta}$—*i.e.* '$\mathbf{\Theta} = \boldsymbol{\theta}$'—has a belief level given by $\mathrm{p}(\boldsymbol{\theta} \mid \mathbf{d}, \mathcal{I})$ once $\mathbf{d}$ is observed. This is the *a posteriori (AP)* pdf of $\mathbf{\Theta}$ [1–3, 5].

Bayes' Rule inverts the direct probability (3) to yield the required AP pdf:

$$\mathrm{p}(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I}) \propto \mathrm{p}(\mathrm{d} \mid \boldsymbol{\theta}, \mathcal{I}) \mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I}). \tag{4}$$

$\mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I})$ is the *prior* on $\underline{\Theta}$. Data-dominated (*i.e.* objective) inference may be achieved using uniform or diffuse priors, as discussed in Section 1. Then, $\mathrm{p}(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I}) \stackrel{\sim}{\propto} l(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I})$ from (4), but this fails to *regularize* the inference, as discussed next.

### 2.1. Prior Regularization

Normalizability of $\mathrm{p}_E(\cdot)$ (2) implies that $\mathrm{p}_E(\mathrm{e} \mid \mathcal{I}) = f(\|\mathrm{e}\|)$, where $f(\cdot)$ is monotonically decreasing and $\|\mathrm{e}\|$ denotes some norm of e. For example, the (complex) Gaussian noise pdf—to be explored in this paper—is $\mathrm{p}_E(\mathrm{e} \mid \mathcal{I}) \propto \exp\left(-\mathrm{e}^H \boldsymbol{\Sigma}^{-1} \mathrm{e}\right)$, where $\mathrm{e}^H \boldsymbol{\Sigma}^{-1} \mathrm{e}$ is a Mahalanobis norm. The Maximum Likelihood (ML) estimate, from (3), is given by $\boldsymbol{\theta}_{\mathrm{ML}} = \arg . \min_{\boldsymbol{\theta}} \|\mathrm{d} - \mathrm{g}(\boldsymbol{\theta})\|$. The criterion may overfit d with $\mathrm{g}(\cdot)$, particularly if $k \not\ll N$. This manifests itself as thresholds in estimation and excessive order in model selection [5, 9]. Prior regularization can be accomplished via an appropriate non-uniform prior, to yield a Maximum *a Posteriori* (MAP) criterion of the type

$$\boldsymbol{\theta}_{\mathrm{MAP}} = \arg . \min_{\boldsymbol{\theta}} \left[\|\mathrm{d} - \mathrm{g}(\boldsymbol{\theta})\| + \beta(\boldsymbol{\theta})\right], \tag{5}$$

where $\beta(\boldsymbol{\theta})$ is introduced via the prior (4). In tasks such as image segmentation [10], for example, Markovian priors may be adopted to encourage connectivity in the inferred label field. Two principles which yield novel regularizing priors for the signal identification task (1) are now presented.

## 3. TRANSFORMATION INVARIANT INFERENCE

Let $\mathrm{h} : \underline{\Theta} \subset \overline{\mathbb{C}}^k \rightarrow \underline{\eta} \subset \overline{\mathbb{C}}^k \mid \eta = \mathrm{h}(\boldsymbol{\theta})$ be a complex, bijective, holomorphic transformation with continuous first partial derivatives. The gradient matrix of the transformation is denoted by $\nabla_{\boldsymbol{\theta}}(\eta^T)$, whose determinant is assumed to be non-zero, $\forall \boldsymbol{\theta} \in \underline{\Theta}$. Denoting the AP inference for $\Theta$ by $\mathrm{p}(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I})$ (assumed positive $\forall \boldsymbol{\theta}$), then the inference for $\eta$ is [1, 3, 4]

$$\mathrm{q}(\eta \mid \mathrm{d}, \mathcal{I}) = \frac{1}{|\nabla_{\boldsymbol{\theta}}(\eta^T)|^2} \mathrm{p}\left[\mathrm{h}^{-1}(\eta) \mid \mathrm{d}, \mathcal{I}\right], \tag{6}$$

where $\mathrm{h}^{-1}(\cdot)$ denotes the inverse transformation. The determinant term, $|\cdot|$, which is squared since the transformation is *complex*, is the hypervolume element expansion ratio for $\mathrm{h}(\cdot)$, and arises in (6) because of the normalization requirement of probabilities.

One consequence of (6) is that *Bayesian* inferences—as opposed to inferences based on non-measure functions such as the LF [5]—are not, in general, invariant with respect to a transformation. For example, the Maximum *a Posteriori* (MAP) estimate will not satisfy $\eta_{\mathrm{MAP}} = \mathrm{h}(\boldsymbol{\theta}_{\mathrm{MAP}})$, which is desirable. Thus, the MAP estimate of $f$ in the signal model $\sin(2\pi f n)$ will not be consistent with that of $T$ in $\sin(2\pi n/T)$, under the transformation $T = 1/f$. However, the following theorem, based on work by Jeffreys [4], provides a sufficient condition for transformation invariance to be met.

**Theorem 1** *Consider hypothesis $\mathcal{I}$, which explains observations $\mathrm{d} \in \overline{\mathbb{C}}^N$ in terms of parameters $\Theta = \boldsymbol{\theta} \in \underline{\Theta} \subset \overline{\mathbb{C}}^k$, via the LF, $l(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I})$. Let $\eta = \mathrm{h}(\boldsymbol{\theta})$ belong to the transformation class defined above. Then AP inferences are invariant with respect to $\mathrm{h}(\cdot)$ if the following prior is used:*

$$\mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I}) \propto |\mathrm{J}(\boldsymbol{\theta} \mid \mathcal{I})|. \tag{7}$$

*$\mathrm{J}(\boldsymbol{\theta} \mid \mathcal{I})$ is the Fisher information matrix under $\mathcal{I}$, whose elements are*

$$J_{ij}(\boldsymbol{\theta}|\mathcal{I}) = -\mathcal{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \ln l(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I})}{\partial \theta_i \partial \theta_j}\right] \qquad 1 \leq i, j \leq k. \tag{8}$$

□

- A proof is given in [7]. The mild regularity conditions which must be met by $l(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I})$ are given in [5].
- The appropriate invariance prior for *real* parameter transformations is $\mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I}) \propto |\mathrm{J}(\boldsymbol{\theta} \mid \mathcal{I})|^{\frac{1}{2}}$.

### 3.1. Data-Translated Likelihood

If the LF can be expressed in the form

$$l(\boldsymbol{\theta} \mid \mathrm{d}, \mathcal{I}) = q\left[\boldsymbol{\phi}(\boldsymbol{\theta}) - f(\mathrm{d})\right], \tag{9}$$

then it is *data-translated* [1] in the transformed space, $\underline{\Phi}$, of $\boldsymbol{\phi}$, since d influences only the *location*, and not the shape, of the LF. While this property often holds in the single parameter case for asymptotically large samples, it is necessary to weaken the definition in the multiparameter case to one of *hypervolume invariance* of the LF. It has been shown [1] that the prior (7) is consistent with the assignment of a *uniform* prior in the space, $\underline{\Phi}$, where this hypervolume invariance is achieved. This confers another justification for choosing (7), though it is important to recognize the weakness of this invariance constraint [4].

### 3.2. Prior Stochastic Independence

If $\Theta = (\boldsymbol{\psi}^T, \boldsymbol{\phi}^T)^T$, and stochastic independence between $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ is to be imposed *a priori*, then the prior factorization

$$\mathrm{p}(\boldsymbol{\theta} \mid \mathrm{k}, \mathrm{l}, \mathcal{I}) \propto |\mathrm{J}_1(\boldsymbol{\psi} \mid \mathrm{k}, \mathcal{I})||\mathrm{J}_2(\boldsymbol{\phi} \mid \mathrm{l}, \mathcal{I})| \tag{10}$$

secures AP invariance under the decoupled transformations $\eta_1 = \mathrm{h}_1(\boldsymbol{\psi})$, $\eta_2 = \mathrm{h}_2(\boldsymbol{\phi})$, where $\mathrm{J}_1(\boldsymbol{\psi} \mid \mathrm{k}, \mathcal{I})$ is calculated from $l(\boldsymbol{\psi} \mid \boldsymbol{\phi} = \mathrm{k}, \mathrm{d}, \mathcal{I})$ via (8), and $\mathrm{J}_2(\boldsymbol{\phi} \mid \mathrm{l}, \mathcal{I})$ is calculated from $l(\boldsymbol{\phi} \mid \boldsymbol{\psi} = \mathrm{l}, \mathrm{d}, \mathcal{I})$. k and l are the necessary *hyperparameters* of the prior [5].

## 4. INVARIANCE PRIOR FOR A WIDE SIGNAL CLASS

Consider the hypothesis, $\mathcal{I}$, which analyzes observations, $d[n]$, in terms of $m$ basis functions, $G_k[n,\boldsymbol{\omega}]$, $\boldsymbol{\omega} \in \underline{\mathbb{R}}^r$:

$$d[n] = \sum_{k=1}^{m} b_k G_k[n,\boldsymbol{\omega}] + e[n], \qquad n = 0,\ldots,N-1. \quad (11)$$

$b_k$ are the linear coefficients, and $e[n]$ are the residuals ('noise'). This rich class [5, 9, 11] embraces a wide range of signal analysis techniques as special cases, including Fourier and wavelet analyses, etc. Gathering the $N$ observations into vector-matrix form, then

$$\mathbf{d} = \mathbf{s} + \mathbf{e} = \mathbf{G}(\boldsymbol{\omega})\mathbf{b} + \mathbf{e}, \quad (12)$$

where $\mathbf{G}(\boldsymbol{\omega}) \in \underline{\mathbb{C}}^{N\times m}$ with $(n,k)$th element $G_k[n,\boldsymbol{\omega}]$, and $\mathbf{e} \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$, being complex, zero-mean, Gaussian noise with known covariance matrix $\boldsymbol{\Sigma}$ (see Section 2.1). From (3,12):

$$l(\boldsymbol{\omega},\mathbf{b} \mid \mathbf{d},\boldsymbol{\Sigma},\mathcal{I}) \propto \exp\left[-(\mathbf{d}-\mathbf{G}\mathbf{b})^H\boldsymbol{\Sigma}^{-1}(\mathbf{d}-\mathbf{G}\mathbf{b})\right], \quad (13)$$

where $\mathbf{G} \equiv \mathbf{G}(\boldsymbol{\omega})$ for convenience. From (7,8,13), matrix calculus yields the following invariance prior for the signal parameter space [5]:

$$\mathrm{p}(\boldsymbol{\omega},\mathbf{b}|\boldsymbol{\Sigma},\mathcal{I}) \propto \left|\Re\begin{bmatrix} \mathbf{b}^H\frac{\partial\mathbf{G}^H}{\partial\omega_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\mathbf{G}}{\partial\omega_j}\mathbf{b} & \mathbf{b}^H\frac{\partial\mathbf{G}^H}{\partial\omega_i}\boldsymbol{\Sigma}^{-1}\mathbf{G} & j\mathbf{b}^H\frac{\partial\mathbf{G}^H}{\partial\omega_i}\boldsymbol{\Sigma}^{-1}\mathbf{G} \\ \mathbf{G}^H\boldsymbol{\Sigma}^{-1}\frac{\partial\mathbf{G}}{\partial\omega_j}\mathbf{b} & \mathbf{G}^H\boldsymbol{\Sigma}^{-1}\mathbf{G} & j\mathbf{G}^H\boldsymbol{\Sigma}^{-1}\mathbf{G} \\ -j\mathbf{G}^H\boldsymbol{\Sigma}^{-1}\frac{\partial\mathbf{G}}{\partial\omega_j}\mathbf{b} & -j\mathbf{G}^H\boldsymbol{\Sigma}^{-1}\mathbf{G} & \mathbf{G}^H\boldsymbol{\Sigma}^{-1}\mathbf{G} \end{bmatrix}\right|^{\frac{1}{2}}. \quad (14)$$

Certain submatrices above are represented by their $(i,j)$th element, $i$th row or $j$th column as appropriate, for notational convenience. Techniques for marginalizing out $\boldsymbol{\Sigma}$, when it is unknown *a priori*, are described in [1, 5]. Useful special cases include the following:

(i) If $\boldsymbol{\omega}$ and $\mathbf{b}$ are independent *a priori*, then, from (10) [5]:

$$\mathrm{p}(\boldsymbol{\omega},\mathbf{b} \mid \boldsymbol{\Sigma},\mathbf{k},\mathcal{I}) \propto \left|\Re\left[\mathbf{k}^H\frac{\partial\mathbf{G}^H}{\partial\omega_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\mathbf{G}}{\partial\omega_j}\mathbf{k}\right]\right|^{\frac{1}{2}}, \quad (15)$$

where the information matrix on the right is denoted by its $(i,j)$th element.

(ii) If *all* $\omega_j$ and $b_k$ are independent, then, from (10) [5]:

$$\mathrm{p}(\boldsymbol{\omega},\mathbf{b} \mid \boldsymbol{\Sigma},\mathbf{k},\mathbf{l},\mathcal{I}) \propto \left[\prod_{i=1}^{r}\mathbf{k}^H\frac{\partial\mathbf{G}_i^H}{\partial\omega_i}\boldsymbol{\Sigma}^{-1}\frac{\partial\mathbf{G}_i}{\partial\omega_i}\mathbf{k}\right]^{\frac{1}{2}}. \quad (16)$$

$\mathbf{G}_i$ denotes $\mathbf{G}$ as a function of $\omega_i$ when *all* other $\omega_j$ are held constant at scalar values which are gathered together in the hyperparameter vector, $\mathbf{l}$, on the lefthand side. Hence, (16) is a product of $r$ scalar functions parameterized successively by $\omega_i$, achieving the necessary *a priori* independence.

## 5. AN ENTROPIC PRIOR FOR SIGNAL IDENTIFICATION

The expected information (*i.e. entropy*)—in the Shannon sense—gained upon observing $\Theta \sim \mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I})$ is given by

$$\mathcal{H}_{\underline{\Theta}} = -\int_{\underline{\Theta}} \mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I}) \ln[\mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I})]\, d\boldsymbol{\theta} \quad (17)$$

(assuming Lebesgue measure on $\underline{\Theta}$ [3, 5]). An intuitively appealing prior assignment strategy is to maximize (17) with respect to $\mathrm{p}(\boldsymbol{\theta} \mid \mathcal{I})$, subject to any testable constraints on $\Theta$. This yields the *minimum information prior* [2, 5–7]. A novel case is now described for the model structure (12).

From (12), the mean square value of the signal over the observation window is $\mathcal{T}(\boldsymbol{\omega},\mathbf{b}) = \frac{1}{N}\mathbf{b}^H\mathbf{G}^H\mathbf{G}\mathbf{b}$, with ensemble average over $\underline{\Theta}$ given by

$$\mathcal{E}[\mathcal{T}(\boldsymbol{\omega},\mathbf{b})] = \gamma^2. \quad (18)$$

$\mathcal{E}[\cdot]$ is with respect to the prior $\mathrm{p}(\boldsymbol{\omega},\mathbf{b} \mid \mathcal{I})$, and $\gamma^2$, the expected square value (*i.e.* 'power'), is testable via an ergodicity assumption. Lagrangian optimization of (17), constrained by (18), yields a density of the kind [5]

$$p(\boldsymbol{\omega},\mathbf{b} \mid \alpha,\mathcal{I}) \propto \exp\left[-\alpha\mathbf{b}^H\mathbf{G}^H\mathbf{G}\mathbf{b}\right], \quad (19)$$

where $\alpha$ is a Lagrange multiplier. Standard manipulations yield $\alpha = m/N\gamma^2$ [5].

- It has been shown [5, 11] that (19) is similar to the *Ockham Prior* for the signal structure (12); *i.e.* it favours objectively 'simpler' models, thereby regularizing the AP inference for $\boldsymbol{\omega}$ and $\mathbf{b}$, in the sense explained by (5);
- (19) may be generalized to the case where there is testable information concerning signal autocorrelations to lag $N-1$, *i.e.*

$$\mathcal{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{C}. \quad (20)$$

$\mathbf{s} = \mathbf{G}\mathbf{b}$ (12) and $\mathbf{C} \in \underline{\mathbb{C}}^{N\times N}$ is the known Hermitian correlation matrix for the signal. Maximization of (17), constrained by (20), yields the following entropic prior [5]:

$$\mathrm{p}(\boldsymbol{\omega},\mathbf{b} \mid \mathbf{C},\mathcal{I}) \propto \exp\left[-\mathbf{b}^H\mathbf{G}^H\mathbf{C}^{-1}\mathbf{G}\mathbf{b}\right]. \quad (21)$$

- Prior independence is not engendered by the priors (19,21). If the second moments of $\mathbf{b}$ and $\boldsymbol{\omega}$ are available independently *a priori*, then conventional independent Gaussian priors are obtained on the spaces, $\underline{\mathbf{B}}$ and $\underline{\mathbf{W}}$, of $\mathbf{b}$ and $\boldsymbol{\omega}$ respectively, using the procedure above [5].

## 6. SUPERRESOLUTION OF CLOSELY SPACED FREQUENCIES

Consider the classic problem of inferring the difference frequency, $\omega$, in the following signal model:

$$d[n] = e^{j\omega_1 n}\left[b_1 + b_2 e^{j\omega n}\right] + e[n], \qquad n = 0, \ldots, N-1, \tag{22}$$

where $E[n] = e[n] \sim \mathcal{N}(0, \sigma^2)$, $\forall\ n$. $\omega_1$ is known *a priori* and $b_1$, $b_2$ and $\omega$ are to be inferred. In the sequel, *marginal a priori* densities for $\omega$,

$$\mathrm{p}(\omega \mid \mathbf{x}, \mathcal{I}) = \int_{\underline{\mathrm{B}}} \mathrm{p}(\omega, \mathbf{b} \mid \mathbf{x}, \mathcal{I}) d\mathbf{b},$$

are compared, where $\mathbf{x}$ denotes any necessary hyperparameters, and the integrand on the right is the joint parameter prior which will be assigned using the strategies outlined in this paper.

### 6.1. Invariance Prior

Assuming $\omega$ and $\mathbf{b}$ are independent *a priori*, then, from (15) or (16), $\mathrm{p}(\omega \mid \sigma, \mathbf{k}, \mathcal{I}) \propto \mathrm{p}(\omega, \mathbf{b} \mid \sigma, \mathbf{k}, \mathcal{I})$ (assuming a range constraint on $\mathbf{b}$). Substituting (22) (via (11) and (12)) into (15) or (16), noting that $\boldsymbol{\omega} = \omega$ (a scalar) in this case, and letting $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_N$ (*i.e.* diagonal), then $\mathrm{p}(\omega \mid \sigma, \mathbf{k}, \mathcal{I}) \propto$ const. (see Fig. 1). In the Figure, one *DFT bin* is defined to be $2\pi/N$ rads/sample. If independence is relaxed *a priori*, then (14) must be employed, yielding a prior which is also illustrated in Fig. 1. It behaves similarly to the one proposed in [12], being approximately uniform for $\omega > 1$ DFT bin, and small for $\omega < 1$ bin, thereby eliminating the possibility of model rank deficiency *a priori*.

### 6.2. Entropic Prior

From (19), the marginal entropic prior for $\omega$ is [5]

$$\begin{aligned}\mathrm{p}(\omega \mid \alpha, \mathcal{I}) &= \int_{\underline{\mathrm{B}}} \mathrm{p}(\omega, \mathbf{b} \mid \alpha, \mathcal{I}) d\mathbf{b} \propto |\mathbf{G}^H \mathbf{G}|^{-1} \\ &= \left(N^2 - \sin^2\left(\frac{N\omega}{2}\right) / \sin^2\left(\frac{\omega}{2}\right)\right)^{-1}, \end{aligned} \tag{23}$$

assuming $\underline{\mathrm{B}} = \underline{\mathrm{C}}^m$. This is also plotted in Fig. 1. As a regularizing prior, it outperforms the others; *i.e.* it is large for $\omega < 1$ bin, and has minima at integer DFT bins, $\omega = k2\pi/N$, $k \in \underline{\mathrm{Z}}^+$ (*i.e.* where the basis functions of (22) are orthogonal [5]). As such, this prior engenders Ockham characteristics in the inference: if data support for the hypothesis (22) is insufficient, (23) dominates the AP inference, $\mathrm{p}(\omega \mid \mathbf{d}, \sigma, \mathcal{I})$, because of the singularity at $\omega = 0$, and this encourages the *rejection* of the hypothesis. If data support is sufficient, the LF will dominate the prior (23), and good estimates will result. As mentioned earlier, (19) and (23) have be deduced, from fundamentals, as the Ockham Prior, whose excellent regularization properties, in both Estimation and Model Selection, have been explored in simulations [5, 9, 11]. Robust, threshold-free inferences are possible over a wide range of Signal-to-Noise Ratios (SNRs) and observation window lengths ($N$).

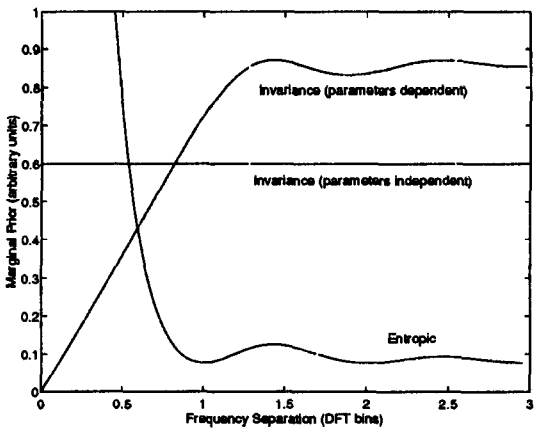


Figure 1: Invariance and Entropic Priors for $\omega$ in the 2-Cis Model (22).

### References


[1] G. E. P. Box and G. C. Tiao. *Bayesian Inference in Statistical Analysis.* Addison-Wesley, 1973.

[2] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 2nd edition, 1985.

[3] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* Wiley, 1994.

[4] H. Jeffreys. *Theory of Probability.* Oxford Univ. Press, 3rd edition, 1961.

[5] A. P. Quinn. *Bayesian Point Inference in Signal Processing.* PhD thesis, Cambridge University Engineering Dept., 1992.

[6] R. D. Rosenkrantz, editor. *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics.* D. Reidel, Dordrecht-Holland, 1983.

[7] A. Zellner. *An Introduction to Bayesian Inference in Econometrics.* John Wiley and Sons, 1971.

[8] G. L. Bretthorst. *Bayesian Spectrum Analysis and Parameter Estimation.* Springer-Verlag, 1989.

[9] A. P. Quinn. A consistent, numerically efficient Bayesian framework for combining the selection, detection and estimation tasks in model-based signal processing. In *Proc. IEEE Int. Conf. on Acoust., Sp. and Sig. Proc.*, Minneapolis, 1993.

[10] O. Schwartz and A. Quinn. Fast and accurate texture-based image segmentation. In *Proc. 3rd IEEE Int. Conf. on Image Proc.*, Lausanne, 1996.

[11] A. Quinn. A general complexity measure for signal identification using Bayesian inference. In *Proc. IEEE Workshop on Nonlinear Sig. and Image Process. (NSIP '95)*, Greece, 1995.

[12] P. M. Djurić and H.-T. Li. Bayesian spectrum estimation of harmonic signals. *IEEE Sig. Proc. Letts.*, 2(11), 1995.