Dirichlet Distribution under Moment Constraints

Thomas A. Mazzuchi
George Washington University, Dept. of Engineering Management and Systems Engineering
Washington, DC 20052, USA
mazzuchi@seas.gwu.edu

Ehsan S. Soofi
University of Wisconsin Milwaukee, School of Business Administration
Milwaukee, WI 53201, USA
esoofi@uwm.edu

Refik Soyer
George Washington University, Dept. of Management Science
Washington, DC 20052, USA
soyer@gwu.edu

1. Introduction

Recently, Mazzuchi, Soofi and Soyer (2000) [MSS (2000)] have introduced an entropy-based Bayesian modeling approach referred to as the Maximum Entropy Dirichlet (MED) procedure. The MED procedure combines the maximum entropy (ME) characterization of the parametric family with a Dirichlet process prior for the unknown data-generating model. In so doing, it uses a ME model as the prior expected distribution in the Dirichlet process prior. The MED procedure provides a Bayesian information index of fit for making inferences about the suitability of a parametric model based on observed data. The procedure provides prior and posterior distributions of a Kullback-Leibler information function in addition to priors and posteriors for various model parameters of interest. The MED procedure is a Monte Carlo based method that requires simulation from Dirichlet distribution under moment constraints implied by the ME characterization. As illustrated in Mazzuchi, Soofi, Soyer and Retzer (2000) [MSSR (2000)] the MED procedure can also be used for modeling categorical data. In what follows, we will give an overview of the Monte Carlo method and the use of the moment constraints in the implementation of the MED procedure.

The entropy of a random variable X with a probability distribution F is defined by

$$H(f) \equiv H[f(x)] = -\int \log f(x) dF(x) \tag{1}$$

where f is the probability density (mass) function of F.

Consider the moment class distributions

$$\Omega_{\boldsymbol{\theta}} = \{f(x|\boldsymbol{\theta}) : E_f[T_j(X)|\boldsymbol{\theta}] = \theta_j,\ j = 0, 1, \cdots, J\}, \tag{2}$$

where $T_j(X)$ are integrable functions with respect to dF and $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)$ is a vector of moment parameters. The maximum entropy (ME) model in the moment class (2), if exists, has the density in the following exponential form:

$$f^*(x|\boldsymbol{\theta}) = C(\boldsymbol{\eta})e^{\eta_1 T_1(x)+\cdots+\eta_J T_J(x)},$$

where the model parameters $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_J)$ are the Lagrange multipliers given by $\eta_j = \eta_j(\boldsymbol{\theta}),\ j = 1, \cdots, J$, and $C(\boldsymbol{\eta}) = \eta_0^{-1}$ is the normalizing constant. The use of a parametric model therefore identifies the moment class $\Omega_{\boldsymbol{\theta}}$.

The closeness of any $f \in \Omega_{\boldsymbol{\theta}}$ to $f^* \in \Omega_{\boldsymbol{\theta}}$ is measured by the Kullback-Leibler (KL) discrimination information function

$$K(f : f^*; \boldsymbol{\theta}) = \int \log \frac{f(x)}{f^*(x)} dF(x) \tag{3}$$

$$= H[f^*(x|\boldsymbol{\theta})] - H[f(x|\boldsymbol{\theta})]; \tag{4}$$

reference for (4) is given in MSS (2000). By making inference on $K(f : f^*; \boldsymbol{\theta})$, we can assess the appropriateness of an ME model to describe unknown data generating distribution F.

2. MED Procedure: Continuous Case

Inference on $K(f : f^*; \boldsymbol{\theta})$ can be obtained by making inference on the entropies $H[f^*(x|\boldsymbol{\theta})]$ and $H[f(x|\boldsymbol{\theta})]$ in (4). The maximum entropy $H[f^*(x|\boldsymbol{\theta})]$ is a parametric function, $H[f^*(x|\boldsymbol{\theta})] = -\log C(\boldsymbol{\eta}) - \eta_1\theta_1 - \cdots - \eta_J\theta_J$. The entropy of the unknown distribution $H[f(x|\boldsymbol{\theta})]$, however, must be estimated nonparametrically. The inference is subject to the moment constraints (2).

The MED procedure uses a Bayesian approach by specifying a Dirichlet process prior to describe uncertainty about the unknown distribution function F with density f as follows.

$$F|F^*, \mathcal{B} \sim \mathcal{D}(F^*, \mathcal{B}),$$

where F^* is the prior best guess for F and $\mathcal{B}$ is the strength of belief parameter. In the MED procedure, F^* is the cumulative distribution function of the ME density $f^*(x|\boldsymbol{\theta}) \in \Omega_{\boldsymbol{\theta}}$ and this choice defines the MED prior for F.

For any given partition of the real line, $-\infty < \xi_1 < \xi_2 < \cdots < \xi_q \leq \infty$, a quantized entropy is defined as

$$H_q(F) = -\sum_{k=1}^{q} \Delta F_k \log \frac{\Delta F_k}{\Delta \xi_k},$$

where $\Delta\xi_k = \xi_k - \xi_{k-1}$ and $\Delta F_k = F_k - F_{k-1},\ F_k = F(\xi_k),\ k = 1, \cdots, q$, with $\xi_0 = sup\{x : F(x) = 0\} \geq -\infty$. In MSS (2000), a specific partition has been constructed using Dirichlet tessalations. Alternative partitions and quantized entropy are discussed in Mazzuchi, Soofi, and Soyer (2001). As discussed in MSS (2000), the posterior distribution of F based on a complete sample from F is also a Dirichlet process with the parameters updated accordingly. Note that for a given partition, distribution of $\boldsymbol{F}$ induces a distribution for the quantized entropy $H_q(F)$. This distribution must be obtained via simulation.

Once the distribution of $H_q(F)$ is simulated the distribution of the the information index $K(f : f^*; \boldsymbol{\theta})$ can be simulated by obtaining the distribution of $H(f^*)$. In so doing, we must ensure the positivity of $K(f : f^*; \boldsymbol{\theta})$. We therefore derive the ME model subject to the moments of the quantized distribution $\boldsymbol{F}$,

$$\theta_{q,j} = \sum_{i=1}^{q} T_j(\bar{\xi}_i)(\Delta F_i), \qquad j = 1, \cdots, J, \tag{5}$$

where $\bar{\xi}_i = (\xi_i + \xi_{i-1})/2$. The quantized moments (5) are approximations of the moments in (2); i.e., $\theta_{q,j} = \theta_j(\boldsymbol{F}) \approx \theta_j = \theta_j(F)$. Thus, for the given partition, for each realization from the Dirichlet process we obtain the moments of the quantized distribution and use these as the moments of the ME model. As a result, we can simulate the distribution of the entropy of the ME model and obtain the distributions of the *KL function.* As pointed out in MSS (2000), we can also simulate the MED distributions of the model parameters under these moment constraints.

3. MED Procedure: Categorical Case

In the categorical data case, the moment class (2) is a set $\mathcal{P}_\theta$ of probability vectors over L categories $\boldsymbol{\pi} = (\pi_1, \cdots, \pi_L)'$ and T_j is usually a vector of covariates or cell indicator functions. The moment *information constraints* may be written as $\boldsymbol{C\pi} = \boldsymbol{\theta}$, where $\boldsymbol{C}$ is the constraint or the *design matrix*, and $\boldsymbol{\theta} = (1, \theta_1, \cdots, \theta_J)'$ is the vector of *moment values.* The rank of $\boldsymbol{C}$ is $J + 1 < L$.

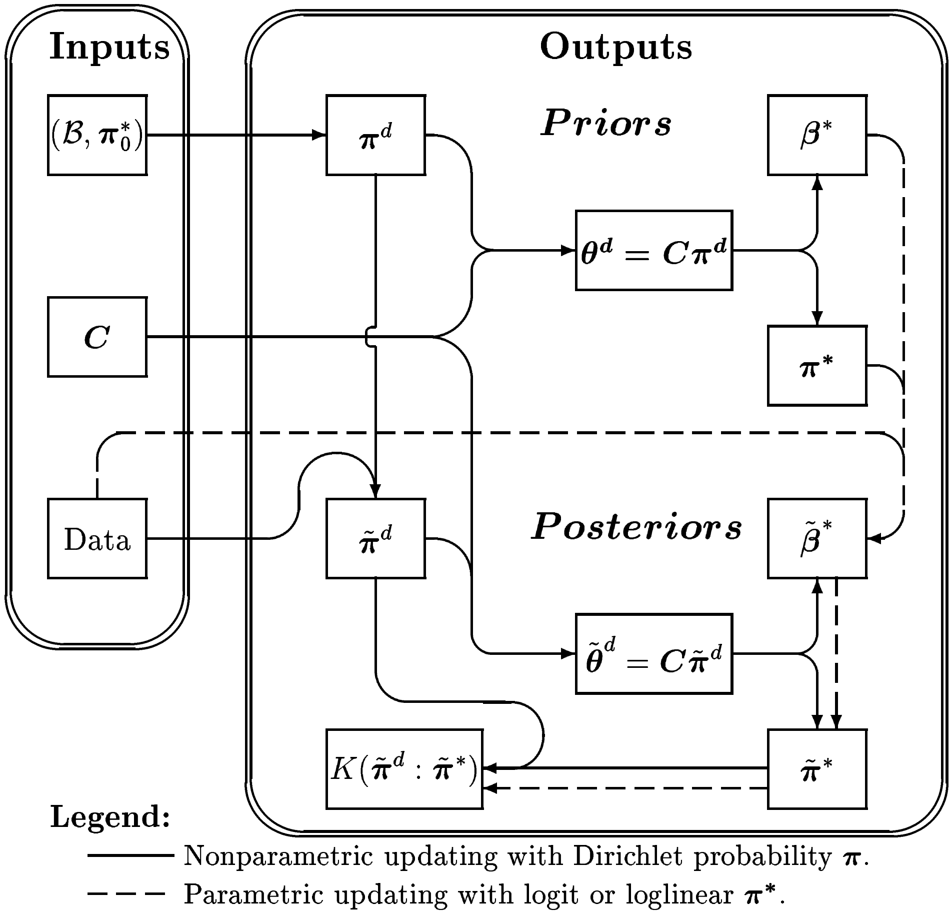
The ME solution is a log-linear model or in a logit form

$$\pi_\ell^* = \frac{e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{c}_\ell}}{\sum_{h=1}^{L} e^{\boldsymbol{\beta}^{*\prime}\boldsymbol{c}_h}}, \tag{6}$$

which is commonly used for modeling $\boldsymbol{\pi}$; here $\boldsymbol{c_\ell} = (C_{1\ell}, \cdots, C_{J\ell})'$ and $\boldsymbol{\beta'} = (\beta_1, \cdots, \beta_J)$ is a vector of Lagrange multipliers (logit parameters).

The flow chart of the MED procedure is shown in Figure 1. As discussed in MSSR (2000), the prior for the $\boldsymbol{\pi} \in \mathcal{P}_\theta$ is a Dirichlet distribution, $\boldsymbol{\pi} \sim \mathcal{D}(\mathcal{B}, \boldsymbol{\pi}^*)$, where the prior expected distribution $\boldsymbol{\pi}^* = \boldsymbol{E(\pi)}$ is the the ME distribution in $\mathcal{P}_\theta$. The probability distributions for the ME model $\boldsymbol{\pi}^*$ and its parameters $\boldsymbol{\beta}^*$ are obtained via the *Bayesian Internal Constraint Problem (BICP).* This is done by obtaining the moments from the Dirichlet probabilities. For

Figure 1. Flow chart of MED Inputs and Outputs.



each realization of the Dirichlet probabilities $\boldsymbol{\pi}^d$, we set the BICP constraints:

$$\boldsymbol{C}\boldsymbol{\pi} = \boldsymbol{\theta}^d = \boldsymbol{C}\boldsymbol{\pi}^d. \tag{7}$$

Then the MDI procedure using the constraint (7) gives a unique $\boldsymbol{\pi}^{*d}$ and $\boldsymbol{\beta}^{*d}$. Therefore, the MED distribution for $\boldsymbol{\pi}$ induces distributions for the model $\boldsymbol{\pi}^*$ and parameter $\boldsymbol{\beta}^*$ in (6):

$$\boldsymbol{\pi}^{*d} = \boldsymbol{\pi}^*(\boldsymbol{\pi}^d), \quad \text{and} \quad \boldsymbol{\beta}^{*d} = \boldsymbol{\beta}^*(\boldsymbol{\pi}^d).$$

Once the data is observed, the posterior distribution of $\boldsymbol{\pi}$ is obtained as a Dirichlet with parameters updated accordingly. The distribution of the KL discrimination information function is obtained directly from (3).

REFERENCES

Mazzuchi, T. A., Soofi, E.S. and Soyer, R. (2000). Computations of Maximum Entropy Dirichlet for Modeling Lifetime Data, *Computational Statistics and Data Analysis*, **32**, 361-378.

Mazzuchi, T. A., Soofi, E.S., Soyer, R. and Retzer, J. J. (2000). Maximum Entropy Dirichlet Modeling of Consumer Choice, *Proc. of ASA Sec. Bayesian Statist. Sci.*, 56-61.

Mazzuchi, T. A., Soofi, E.S. and Soyer, R. (2001). Bayesian Estimation of Entropy and Information Index of Fit. *Technical Report.*