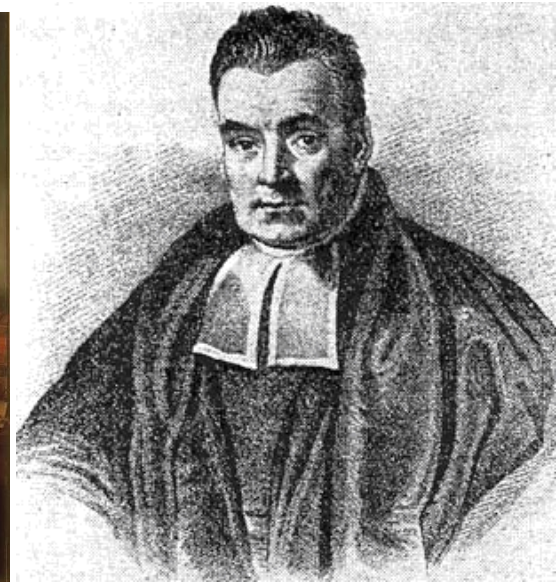


Comparison Between Bayesian and Maximum Entropy Analysis of Flow Networks



Maximum Entropy

- The Maximum Entropy (MaxEnt) method is a methodology to assign or update probability distributions to describe systems which are not completely specified
- The probability distribution inferred by MaxEnt only contains the information given in its constraints and a prior distribution
- The constraints can incorporate physical laws such as conservation equations
- The inferred distribution can be interpreted as the most probable distribution which represents the system

MaxEnt Methodology

- Define a probability measure over the uncertainties of interest
- Construct a relative entropy function
- Identify constraints
- Maximize the entropy function subject to constraints and prior
- If desired extract statistical moments of quantities of interest

Combinatorial Basis of the MaxEnt Method

- The Combinatorial approach results in the MaxEnt distribution being the one in which the states of a system can be realised in the greatest number of ways.
- The Boltzmann entropy is given by:

$$S = \frac{\ln(\mathbb{P})}{N} \quad \text{where } \mathbb{P} = N! \prod_{i=1}^s \frac{q_i^{n_i}}{n_i!} \quad \text{and } p_i = \frac{n_i}{N}$$

$$N \rightarrow \infty \quad S = -\sum_{i=1}^s p_i \ln\left(\frac{p_i}{q_i}\right)$$

$$p_i \rightarrow p(x)dx, \text{ and } q_i \rightarrow q(x)dx, \quad S = \int_{\Omega} p(x) \ln\left(\frac{p(x)}{q(x)}\right) dx$$

The Axiomatic Basis of the MaxEnt Method

- **Locality** - Constraints giving information about a sub-domain and no information about what is happening outside the sub-domain should only update the sub-domain where new information is provided. Also if no constraints are applied no changes should be made to the current belief.
- **Coordinate invariance** - The inference should give the same outcome irrespective of the coordinate system used to conduct the inference.
- **Subsystem independence** - If a system is formed from subsystems which are believed to be independent it should not matter whether the inference procedure analyses the subsystems separately or jointly.

Bayesian Inference

- Bayesian inference is a method in which Bayes theorem is used to update a prior to a posterior using data.
- To use Bayes theorem the prior and likelihood functions need to be chosen before the data is analysed.
- To analyse the data, a set of data values are evaluated in the likelihood function(s), the likelihood function(s) are multiplied by the prior, then normalized to obtain the posterior.
- This process can be repeated for each data set by using the posterior as the prior for the next data set.

Probability Rules

Product Rule

$$p(x, y) = p(x | y)p(y) = p(y | x)p(x)$$

Marginalization

$$p(y) = \int_{\Omega} p(x, y) dx$$

$$p(x) = \int_{\Omega} p(x, y) dy$$

Bayes Rule

$$p(x | y) = \frac{p(y | x)p(x)}{p(y)}$$

Bayesian Analysis of Flow Networks

Interested in flow rates on the network

$$\Psi = \begin{bmatrix} \Theta \\ Q \end{bmatrix}$$

As networks can be represented differently a subset of flow rates are inferred using Bayes theorem

$$p(\mathbf{X}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{X})q(\mathbf{X})}{\int_{\Omega} \dots \int p(\mathbf{y}|\mathbf{X})q(\mathbf{X})d\mathbf{X}}$$

The remaining flows can be found from the inferred flow rates

$$\bar{\mathbf{X}} = \mathbf{V}\mathbf{X}$$

$$\mathbf{V} = -\mathbf{A}_{i \in \mathcal{V}, j \in \mathcal{D}}^{-1} \mathbf{A}_{i \in \mathcal{V}, j \in \mathcal{E}}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{C} \\ \mathbf{W} \text{diag}(\mathbf{K}) \\ \mathbf{F} \\ \mathbf{T} \text{diag}(\mathbf{K}) \end{bmatrix}$$

Prior:

$$q(\mathbf{X}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{X} - \mathbf{m})^{\top} \boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{m})\right)}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}}$$

C=coefficient matrix for conservation of mass

W=coefficient matrix for loop laws

F=coefficient matrix for observed flow rates

T=coefficient matrix for observed potential differences

K=vector of flow rate resistances

y=vector of observed data

Sets:

\mathcal{V} =indices of equations used to find $\bar{\mathbf{X}}$ from \mathbf{X}

\mathcal{E} =indices of dimensions of inference

\mathcal{D} =indices of dimensions found from inferred values

Likelihood Functions

Kirchhoff's first law: conservation of mass

$$p(\mathbf{0}|\mathbf{X}) = \delta(\mathbf{0} - (\mathbf{C}_X + \mathbf{C}_{\bar{X}}) \mathbf{X}) \quad -2 \ln(p(\mathbf{0}|\mathbf{X})) \propto \lim_{\Sigma_C \rightarrow 0} (\mathbf{0} - (\mathbf{C}_X + \mathbf{C}_{\bar{X}}) \mathbf{X})^\top \Sigma_C^{-1} (\mathbf{0} - (\mathbf{C}_X + \mathbf{C}_{\bar{X}}) \mathbf{X})$$

$$\mathbf{C}_X = \mathbf{C}_{i \notin \mathcal{V}, j \in \mathcal{E}}$$

Kirchhoff's second law: Loop laws

$$\mathbf{C}_{\bar{X}} = \mathbf{C}_{i \notin \mathcal{V}, j \in \mathcal{D}} \mathbf{V}$$

$$p(\mathbf{0}|\mathbf{X}) = \delta(\mathbf{0} - (\mathbf{W}_X + \mathbf{W}_{\bar{X}}) \mathbf{X}) \quad -2 \ln(p(\mathbf{0}|\mathbf{X})) \propto \lim_{\Sigma_W \rightarrow 0} (\mathbf{0} - (\mathbf{W}_X + \mathbf{W}_{\bar{X}}) \mathbf{X})^\top \Sigma_W^{-1} (\mathbf{0} - (\mathbf{W}_X + \mathbf{W}_{\bar{X}}) \mathbf{X})$$

$$\mathbf{W}_X = \mathbf{W}_{i \notin \mathcal{V}, j \in \mathcal{E}} \text{diag}(\mathbf{K}_{i \in \mathcal{E}})$$

Flow rate observations

$$\mathbf{W}_{\bar{X}} = \mathbf{W}_{i \notin \mathcal{V}, j \in \mathcal{D}} \text{diag}(\mathbf{K}_{i \in \mathcal{D}}) \mathbf{V}$$

$$-2 \ln(p(\mathbf{Y}_F|\mathbf{X})) \propto (\langle \mathbf{Y}_F \rangle - (\mathbf{F}_X + \mathbf{F}_{\bar{X}}) \mathbf{X})^\top \mathbf{N}_F \Sigma_F^{-1} (\langle \mathbf{Y}_F \rangle - (\mathbf{F}_X + \mathbf{F}_{\bar{X}}) \mathbf{X})$$

$$\mathbf{F}_X = \mathbf{F}_{i \notin \mathcal{V}, j \in \mathcal{E}}$$

$$\mathbf{F}_{\bar{X}} = \mathbf{F}_{i \notin \mathcal{V}, j \in \mathcal{D}} \mathbf{V}$$

Potential difference observations

$$-2 \ln(p(\mathbf{Y}_T|\mathbf{X})) \propto (\langle \mathbf{Y}_T \rangle - (\mathbf{T}_X + \mathbf{T}_{\bar{X}}) \mathbf{X})^\top \mathbf{N}_T \Sigma_T^{-1} (\langle \mathbf{Y}_T \rangle - (\mathbf{T}_X + \mathbf{T}_{\bar{X}}) \mathbf{X})$$

$$\mathbf{T}_X = \mathbf{T}_{i \notin \mathcal{V}, j \in \mathcal{E}} \text{diag}(\mathbf{K}_{i \in \mathcal{E}})$$

$$\mathbf{T}_{\bar{X}} = \mathbf{T}_{i \notin \mathcal{V}, j \in \mathcal{D}} \text{diag}(\mathbf{K}_{i \in \mathcal{D}}) \mathbf{V}$$

Posterior

$$-2 \ln(p(\mathbf{X}|\mathbf{y})) \propto \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} - \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{m} - \mathbf{m}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X} + \mathbf{X}^\top \mathbf{O}^\top \mathbf{S}^{-1} \mathbf{O} \mathbf{X} - \mathbf{y}^\top \mathbf{S}^{-1} \mathbf{O} \mathbf{X} - \mathbf{X}^\top \mathbf{O}^\top \mathbf{S}^{-1} \mathbf{y}$$

where

$$\mathbf{O} = \begin{bmatrix} \mathbf{C}_X + \mathbf{C}_X \\ \mathbf{W}_X + \mathbf{W}_X \\ \mathbf{F}_X + \mathbf{F}_X \\ \mathbf{T}_X + \mathbf{T}_X \end{bmatrix}$$

$$\mathbf{S}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}_C^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_W^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & N_F \boldsymbol{\Sigma}_F^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & N_T \boldsymbol{\Sigma}_T^{-1} \end{bmatrix}$$

$$\mathbf{y} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \langle \mathbf{Y}_F \rangle \\ \langle \mathbf{Y}_T \rangle \end{bmatrix}$$

combining like factors

$$-2 \ln(p(\mathbf{X}|\mathbf{y})) \propto \mathbf{X}^\top \left(\boldsymbol{\Sigma}^{-1} + \mathbf{O}^\top \mathbf{S}^{-1} \mathbf{O} \right) \mathbf{X} - \mathbf{X}^\top \left(\boldsymbol{\Sigma}^{-1} \mathbf{m} + \mathbf{O}^\top \mathbf{S}^{-1} \mathbf{y} \right) - \left(\mathbf{m}^\top \boldsymbol{\Sigma}^{-1} + \mathbf{y}^\top \mathbf{S}^{-1} \mathbf{O} \right) \mathbf{X}$$

Completing the square gives

$$-2 \ln(p(\mathbf{X}|\mathbf{y})) \propto (\mathbf{X} - \langle \mathbf{X} \rangle)^\top \boldsymbol{\Sigma}_p^{-1} (\mathbf{X} - \langle \mathbf{X} \rangle)$$

Posterior Mean and Covariance

$$\langle X \rangle = \Sigma_p \left(\Sigma^{-1} m + O^\top S^{-1} y \right)$$

$$\Sigma_p = \left(\Sigma^{-1} + O^\top S^{-1} O \right)^{-1}$$

$$\langle X \rangle = m + \Sigma O^\top \left(S + O \Sigma O^\top \right)^{-1} (y - O m) \quad \Sigma_p = \Sigma - \Sigma O^\top \left(S + O \Sigma O^\top \right)^{-1} O \Sigma$$

Maximum Entropy Analysis of Flow Networks With Soft Constraints

Probability: $p(\mathbf{X})d\mathbf{X} = \text{Prob}(\mathbf{X} \leq \mathbf{Y}_X \leq \mathbf{X} + d\mathbf{X}, \mathbf{Y}_F \leq \mathbf{Y}_{Y_F} \leq \mathbf{Y}_F + d\mathbf{Y}_F, \mathbf{Y}_T \leq \mathbf{Y}_{Y_T} \leq \mathbf{Y}_T + d\mathbf{Y}_T)$

Relative Entropy Function:
$$\mathfrak{H} = - \int_{l_1}^{u_1} \cdots \int_{l_{n+n_0}}^{u_{n+n_0}} p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) \ln \frac{p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T)}{q(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T)} d\mathbf{X} d\mathbf{Y}_F d\mathbf{Y}_T$$

Constraints:

Normalization:
$$1 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_0}}^{u_{n+n_0}} p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) d\mathbf{X} d\mathbf{Y}_F d\mathbf{Y}_T$$

Kirchhoff's first law: conservation of mass
$$0 = (C_X + C_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_0}}^{u_{n+n_0}} \mathbf{X} p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) d\mathbf{X} d\mathbf{Y}_F d\mathbf{Y}_T \right)$$

Kirchhoff's second law: Loop laws
$$0 = (W_X + W_{\bar{X}}) \left(\int_{l_1}^{u_1} \cdots \int_{l_{n+n_0}}^{u_{n+n_0}} \mathbf{X} p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) d\mathbf{X} d\mathbf{Y}_F d\mathbf{Y}_T \right)$$

Flow rate observations
$$0 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_0}}^{u_{n+n_0}} ((F_X + F_{\bar{X}}) \mathbf{X} - \mathbf{Y}_F) p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) d\mathbf{X} d\mathbf{Y}_F d\mathbf{Y}_T$$

Potential difference observations
$$0 = \int_{l_1}^{u_1} \cdots \int_{l_{n+n_0}}^{u_{n+n_0}} ((T_X + T_{\bar{X}}) \mathbf{X} - \mathbf{Y}_T) p(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) d\mathbf{X} d\mathbf{Y}_F d\mathbf{Y}_T$$

Inferred Distribution:

$$p^*(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) = q(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) e^{-\kappa - \alpha(C_X + C_{\bar{X}})\mathbf{X} - \beta(W_X + W_{\bar{X}})\mathbf{X} - \lambda((F_X + F_{\bar{X}})\mathbf{X} - \mathbf{Y}_F) - \eta((T_X + T_{\bar{X}})\mathbf{X} - \mathbf{Y}_T)}$$

Maximum Entropy Prior Including Data Observations and Inferred Distribution

$$q(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T) \propto \exp \left(-\frac{1}{2} \left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y}_F \\ \mathbf{Y}_T \end{bmatrix} - \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_F \\ \mathbf{m}_T \end{bmatrix} \right)^\top \begin{bmatrix} \Sigma^{-1} & 0 & 0 \\ 0 & \Sigma_F^{-1} & 0 \\ 0 & 0 & \Sigma_T^{-1} \end{bmatrix} \left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y}_F \\ \mathbf{Y}_T \end{bmatrix} - \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_F \\ \mathbf{m}_T \end{bmatrix} \right) \right)$$

$$\ln(p^*(\mathbf{X}, \mathbf{Y}_F, \mathbf{Y}_T)) \propto -\frac{1}{2} \left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y}_F \\ \mathbf{Y}_T \end{bmatrix} - \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_F \\ \mathbf{m}_T \end{bmatrix} + \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} \mathbf{C}_X + \mathbf{C}_{\bar{X}} & 0 & 0 \\ \mathbf{W}_X + \mathbf{W}_{\bar{X}} & 0 & 0 \\ \mathbf{F}_X + \mathbf{F}_{\bar{X}} & -I & 0 \\ \mathbf{T}_X + \mathbf{T}_{\bar{X}} & 0 & -I \end{bmatrix}^\top \gamma \right)^\top$$

$$\times \Sigma^{-1} \left(\begin{bmatrix} \mathbf{X} \\ \mathbf{Y}_F \\ \mathbf{Y}_T \end{bmatrix} - \begin{bmatrix} \mathbf{m} \\ \mathbf{m}_F \\ \mathbf{m}_T \end{bmatrix} + \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} \mathbf{C}_X + \mathbf{C}_{\bar{X}} & 0 & 0 \\ \mathbf{W}_X + \mathbf{W}_{\bar{X}} & 0 & 0 \\ \mathbf{F}_X + \mathbf{F}_{\bar{X}} & -I & 0 \\ \mathbf{T}_X + \mathbf{T}_{\bar{X}} & 0 & -I \end{bmatrix}^\top \gamma \right)$$

where $\gamma = [\alpha \quad \beta \quad \lambda \quad \eta]^\top$.

Maximum Entropy Mean and Lagrange Multipliers

$$\begin{bmatrix} \langle X \rangle \\ \langle Y_F \rangle \\ \langle Y_T \rangle \end{bmatrix} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} X \\ Y_F \\ Y_T \end{bmatrix} p(X, Y_F, Y_T) dX dY_F dY_T = \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} - \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^{\top} \gamma$$

$$\gamma = \left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} \begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix}^{\top} \right)^{-1} \left(\begin{bmatrix} C_X + C_{\bar{X}} & 0 & 0 \\ W_X + W_{\bar{X}} & 0 & 0 \\ F_X + F_{\bar{X}} & -I & 0 \\ T_X + T_{\bar{X}} & 0 & -I \end{bmatrix} \begin{bmatrix} m \\ m_F \\ m_T \end{bmatrix} \right)$$

$$\langle X \rangle = m - \Sigma O^{\top} \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} + O \Sigma O^{\top} \right)^{-1} \left(O m - \begin{bmatrix} 0 \\ m_F \\ m_T \end{bmatrix} \right)$$

Bayesian and Maximum Entropy Mean

Bayesian Mean

$$\langle X \rangle = m + \Sigma O^\top (S + O \Sigma O^\top)^{-1} (y - Om)$$

Maximum Entropy Mean

$$\langle X \rangle = m - \Sigma O^\top \left(\begin{bmatrix} 0 & 0 & 0 \\ 0 & \Sigma_F & 0 \\ 0 & 0 & \Sigma_T \end{bmatrix} + O \Sigma O^\top \right)^{-1} \left(Om - \begin{bmatrix} 0 \\ m_F \\ m_T \end{bmatrix} \right)$$

Conclusions

- The MaxEnt and Bayesian methods rest on different theoretical foundations although they are both able to predict flows on networks by updating a prior belief to a posterior with the inclusion of new information in the form of constraints or uncertain data.
- Through deriving a Bayesian method to analyse flow networks and comparing it with the equivalent MaxEnt method, it has been shown that both the MaxEnt and Bayesian methods give the same mean flow rate prediction when normal distribution priors are used but different covariance predictions.

Acknowledgements

- This project was funded by the Australian Research Council Discovery Project DP140104402; the Go8 / DAAD Australia-Germany Joint Research Cooperation Scheme; Region Poitou-Charentes, France; and TUCOROM, Institut PPrime, Poitiers, France.