Notation	Explanation
w_i	Weight of ith grid node
$\boldsymbol{W}$	$\{w_1, \ldots, w_N\}$
$Z^{(k)} = \{i_1^{(k)}, \ldots, i_{n^{(k)}}^{(k)}\}$	Path taken by trip k
$x_1^{(k)}, x_2^{(k)}$	Start and end locations of trip k
$\boldsymbol{Z}^{(k)} = \boldsymbol{Z}(x_1^{(k)}, x_2^{(k)})$	Set of reasonable paths for trip k
$T^{(k)}$	Time taken by trip k

Assume the following model

$$\begin{aligned}
\boldsymbol{W} &\sim p_{\boldsymbol{W}} \\
Z^{(k)}|\boldsymbol{W} &\sim p_{Z|\boldsymbol{W}} && \text{(a distribution over } \boldsymbol{Z}^{(k)}) \\
T^{(k)}|Z^{(k)}, \boldsymbol{W} &\sim \mathcal{N}\left(\sum_{i \in Z^{(k)}} w_i, \sigma^2\right)
\end{aligned}$$

We are interested in the posterior distribution of $\boldsymbol{W}$:

$$\begin{aligned}
p(\boldsymbol{W}|T^{(k)}) &= \frac{p(T^{(k)}|\boldsymbol{W})p(\boldsymbol{W})}{p(T^{(k)})} \\
&= \frac{\left[\sum_{Z^{(k)} \in \boldsymbol{Z}^{(k)}} p(T^{(k)}|Z^{(k)}, \boldsymbol{W})p(Z^{(k)}|\boldsymbol{W})\right] p(\boldsymbol{W})}{p(T^{(k)})}
\end{aligned}$$

If we are only interested in the MAP, then we have

$$\log p(\boldsymbol{W}|T^{(k)}) = \underbrace{\log\left(\sum_{Z^{(k)} \in \boldsymbol{Z}^{(k)}} p(T^{(k)}|Z^{(k)}, \boldsymbol{W})p(Z^{(k)}|\boldsymbol{W})\right)}_{\ell(\boldsymbol{W}|T^{(k)}) \text{ log-likelihood}} + \underbrace{\log p(\boldsymbol{W})}_{\text{log prior}} + \text{const.}$$

Our main challenge here is to compute the log-likelihood, which involves a sum over all of $\boldsymbol{Z}^{(k)}$. If the set of reasonable paths include all paths inside a rectangle of size $n \times m$, then we have $\binom{n+m}{n} \sim \frac{(n+m)^n}{n^n} e^n$ terms, which is clearly intractable. Zhan, et al. solves the issue by computing the 20 shortest path as the reasonable path set and then running softmax.

Can we EM? EM algorithm approximates the log-likelihood via

$$\begin{aligned}
Q(\boldsymbol{W}|\boldsymbol{W}^{(t)}) &= E\left[\log p(T^{(k)}, Z^{(k)}|\boldsymbol{W})|T^{(k)}, \boldsymbol{W}^{(t)}\right] \\
&= -\frac{1}{2\sigma^2} \sum_{Z^{(k)} \in \boldsymbol{Z}^{(k)}} \left(T^{(k)} - \sum_{Z^{(k)}} w_i\right)^2 p(Z^{(k)}|T^{(k)}, \boldsymbol{W}^{(t)}) + \text{const.}
\end{aligned}$$

However, (1) we again have to evaluate a sum of order $|\boldsymbol{Z}|$ and (2) we don't have good estimates of $p(Z^{(k)}|T^{(k)}, \boldsymbol{W}^{(t)})$.

Can we Monte Carlo? We might consider using rejection sampling to approximate

$$E_Z\left[p(T^{(k)}|Z^{(k)},\boldsymbol{W})|\boldsymbol{W}\right] \approx \frac{1}{L}\sum_{j=1}^{L} p(T^{(k)}|\tilde{Z}_j^{(k)},\boldsymbol{W}).$$

The problem here is that the gradient update via `autograd` cannot take into account the gradient incurred *in the sampling process*.

Can we variational inference? Maybe. But (1) mean field doesn't really apply since $Z^{(k)}$ cannot be broken into independent variables naturally and (2) VAE a la Kingma and Welling only works on continuous latent variables.

A Monte Carlo based method We repeatedly sample $\tilde{Z}$ *uniformly* from $\boldsymbol{Z}$ and use the sample mean to approximate summations over $\boldsymbol{Z}$.

$$\begin{aligned}
&\log E_Z\left(p(T^{(k)}|Z^{(k)},\boldsymbol{W})|\boldsymbol{W}\right)\\
&= \log\left[|\boldsymbol{Z}|E_{\tilde{Z}}\left[p(T^{(k)}|\tilde{Z}^{(k)},\boldsymbol{W})p(\tilde{Z}^{(k)}|\boldsymbol{W})\right]\right] && \text{(where } \tilde{Z} \text{ is uniform over } \boldsymbol{Z}\text{)}\\
&= \log E_{\tilde{Z}}\left[p(T^{(k)}|\tilde{Z}^{(k)},\boldsymbol{W})p(\tilde{Z}^{(k)}|\boldsymbol{W})\right] + \text{const.}\\
&\geq E_{\tilde{Z}}\left[\log p(T^{(k)}|\tilde{Z}^{(k)},\boldsymbol{W})\right] + E_{\tilde{Z}}\left[\log p(\tilde{Z}^{(k)}|\boldsymbol{W})\right] && \text{(Jensen; omitting constant)}\\
&= -\frac{1}{2L\sigma^2}\sum_{j=1}^{L}\left[T^{(k)} - \sum_{i\in\tilde{Z}_j^{(k)}} w_i\right]^2 + \frac{1}{L}\sum_{i=1}^{L}\left[-\sum_{i\in\tilde{Z}_j^{(k)}} w_i - \log\sum_{\boldsymbol{Z}}\exp\left(-\sum_{Z^{(k)}} w_i\right)\right]\\
&\qquad\qquad \text{(We parameterize the distribution of } Z^{(k)} \text{ as Categorical with } \tfrac{\exp(-\sum w)}{\sum\exp(-\sum w)}\text{)}\\
&= -\frac{1}{2L\sigma^2}\sum_{j=1}^{L}\left[T^{(k)} - \sum_{i\in\tilde{Z}_j^{(k)}} w_i\right]^2 - \frac{1}{L}\sum_{i=1}^{L}\sum_{i\in\tilde{Z}_j^{(k)}} w_i - \sum_{i=1}^{L}\log\left[|\boldsymbol{Z}|E_{\tilde{Z}}\left[\exp\left(-\sum_{\tilde{Z}} w_i\right)\right]\right]\\
&= -\frac{1}{2L\sigma^2}\sum_{j=1}^{L}\left[T^{(k)} - \sum_{i\in\tilde{Z}_j^{(k)}} w_i\right]^2 - \frac{1}{L}\sum_{i=1}^{L}\sum_{i\in\tilde{Z}_j^{(k)}} w_i - \sum_{i=1}^{L}\underset{i=1,\ldots,L}{\text{logsumexp}}\left(-\sum_{\tilde{Z}_i} w_i\right)
\end{aligned}$$