**Introduction**

EP is a widely used message passing based inference algorithm. 
- **Problem**: Expensive to compute outgoing from incoming messages.
- **Goal**: Speed up computation by a cheap regression function (message operator): 
  
  \[ m_{f \rightarrow i}(v_i) = \text{proj} \left[ f(V) \prod_{j=1}^c m_{V_j \rightarrow f}(v_j) \right] \]

**Expectation Propagation (EP)**

Under an approximation that each factor fully factorizes, an outgoing EP message \( m_{f \rightarrow i}(v_i) \) takes the form

\[
\text{proj}[f(V) \prod_{j=1}^c m_{V_j \rightarrow f}(v_j)] := q_{f \rightarrow i}(v_i) = \arg\min q \in \text{ExpFam} \quad \text{KL} \left[ \frac{r_{f \rightarrow i}}{q} \right]
\]

**(projection onto exponential family)**

**Incoming messages \( i \) \rightarrow outgoing message.**

**Merits:**
- Efficient online update of the operator during inference.
- Uncertainty monitored to invoke new training examples when needed.
- Automatic random feature representation of incoming messages.

**Kernel on Incoming Messages**

Propose to incrementally learn during inference a kernel-based EP message operator (distribution-to-distribution regression)

\[
m_{f \rightarrow i}(v_i) \rightarrow q_{f \rightarrow i}(v_i)
\]

**Message Operator: Bayesian Linear Regression**

- **Input**: \( X = (x_1, \cdots, x_N) \): \( N \) training incoming messages represented as random feature vectors.
- **Output**: \( Y = \{E_x(u_i)\} \cdots \{E_x(u_i)\} \in \mathbb{R}^{2 \times N} \): expected sufficient statistics of outgoing messages.
- **Inexpensive online updates of posterior mean and covariance.**
- **Bayesian regression gives prediction and predictive variance.**
- **If predictive variance > threshold, query the importance sampling oracle.**

**Experiment 1: Uncertainty Estimates**

- **Approx**: \( f(p; z) = \delta[p - \frac{1}{1 + \exp(-z)}] \)
- **Incoming messages**: \( m_{w \rightarrow f} = \mathcal{N}(z_i; \mu, \sigma) \), \( m_{w \rightarrow f} = \text{Beta}(p; \alpha, \beta) \).

**Experiment 2: Real Data**

- Binary logistic regression. Sequentially present 4 real datasets to the operator.
- Diverse distributions of incoming messages.

**Experiment 3: Compound Gamma Factor**

Infer posterior of the precision \( \tau \) of \( x \sim \mathcal{N}(x; 0, \tau^{-1}) \) from observations \( \{z_i\}_{i=1}^N \):

\[
x_2 \sim \text{Gamma}(r_2; s_1, r_1) \\
\tau \sim \text{Gamma}(s_2; r_2, r_2) \\
(\{s_1\}, \{r_1\}) \sim \mathcal{N}(0, \tau^{-1})
\]

**Contact**: wittawat@gatsby.ucl.ac.uk  
**Code download**: http://github.com/wittawatj/kernel-ep  
We thank the Gatsby Charitable Foundation for the financial support.