Suppose we have access to:

$$
\begin{array}{ccccccc}
x_1 & , & y_{11}, \ldots, y_{1m}, & f(y_{11}), \ldots, f(y_{1m}) \\
\vdots & & \vdots \quad \vdots & \vdots \quad \vdots \\
x_T & , & y_{T1}, \ldots, y_{Tm}, & f(y_{T1}), \ldots, f(y_{Tm})
\end{array}
$$

& are interested in the integrals:

$$\int f(y)\, p(y \mid X = x_1)\, dy$$

$$\vdots$$

$$\int f(y)\, p(y \mid X = x_T)\, dy$$

Or even $\int f(y)\, p(y \mid X = x)\, dy$ for arbitrary x!

These need to be estimated simultaneously.

Note that:

$$\int f(y)\, p(y \mid X = x^*)\, dy$$

$$= \int \langle f, k(\cdot, y) \rangle_{\mathcal{H}_y}\, p(y \mid X = x^*)\, dy$$

$$= \langle f, \underbrace{\int k(\cdot, y)\, p(y \mid X = x^*)\, dy}_{\mu_{Y \mid X = x^*}} \rangle_{\mathcal{H}_y}$$

A reasonable strategy would therefore be to estimate $\mu_{Y \mid X = x^*}$ from samples, then use the approximation in the inner product above (i.e doing quadrature is equivalent to approximating a kernel mean).

Note that our data is exactly what is needed for the empirical versions of conditional kernel mean embeddings.

We could even estimate the integral for

$$x^* \notin \{x_1, \dots, x_T\}$$

which is pretty cool!

The main difference between the conditional embedding setting & the quadrature setting is that we observe function values exactly!

$$\hat{\mu}_g^n(x)(\cdot) := k_x(x, x_{1:n})^T \left(k_x(x_{1:n}, x_{1:n}) + n\lambda I_{d_n}\right)^{-1} k_y(\cdot, y_{1:n})$$

- $\mathbb{E}_{Y|X=x^*}[f(Y)]$

  $= \langle f, \hat{\mu}_g^n(x^*) \rangle$

- If we look at WCE over $\mathcal{H}_y$ we only get a difference of conditional kernel mean embeddings.

- Can use Arthur's paper to get rates in noisy setting?

- Functional GP regression! with the same kernel as they do vv-regression.

- posterior variance should be MCMD^2 evaluated at x^*!

$$\hat{\Pi}_{CMC}[f] := k_{x}(x, x_{1:m})^T \left(K_x(x_{1:m}, x_{1:m}) + m\lambda I_m\right)^{-1}$$

$$\int (\cdot) \, f(y_{1:n})$$

$$= \sum w_i \, f(y_i)$$

$$\sum w$$

$$\mathcal{L}(F) = \frac{1}{n} \sum_{i=1}^{n} \| F(x_i)(\cdot) - \mu_q(x_i)(\cdot) \|_{\mathcal{H}_y} + \lambda \| F \|_{\mathcal{H}_{yx}}$$

$$\tilde{\mathcal{L}}(F) = \frac{1}{n} \sum_{i=1}^{n} \| F(x_i)(\cdot) - k_y(y_i, \cdot) \|_{\mathcal{H}_y} + \lambda \| F \|_{\mathcal{H}_{yx}}$$

k_y: Matern Gaussian

$\uparrow \to$ Gaussian.

$$\mu_q(x_i)(\cdot) = \int k_y(\cdot, y) \, q(y \mid X = x_i) \, dy$$

$$\approx \sum w_i^{BQ} k_{\mathcal{Y}}(\cdot, y_i)$$

$$\tilde{\mathcal{L}}(F) = \frac{1}{n} \sum_{i=1}^{n} \| F(x_i)(\cdot) - \underbrace{\frac{1}{m}\sum_{i=1}^{m} k_{\mathcal{Y}}(y_i, \cdot)} \|_{\mathcal{H}_{\mathcal{Y}}}$$

$$\hat{\Pi}_{BQ}[f] = \sum w_i^{BQ} f(x_i)$$

$$\int k(x, x_i)\, \pi(dx)$$